

30 April 1981

New guard at the National Academy

The annual meeting of the National Academy of Sciences in Washington this week is something of a milestone. Among other things, the scientific programme includes a meeting to mark the impending retirement of Dr Philip Handler as president of the academy. From then until 1 July, when Dr Frank Press takes over, he will be something of a lame duck, which does not mean that he will be less outspoken than in the past. (President Eisenhower, one of the most taciturn and even inarticulate of presidents, waited until his lame-duck period before making his unexpected but still memorable attack on the "military-industrial complex".) Inevitably, however, the academy's army of committee members will from now on be guessing at the changes that Press's succession is likely to bring. Handler will have to try harder if he seeks to capture their attention.

He should not feel slighted. Handler has done a distinguished job in his two six-year stints as president of the academy. Neither of his immediate predecessors could have done as well. Dr Detlev W. Bronk's apparently randomly directed ebullience may have suited the spirit of the late fifties and early sixties, but would have landed him perpetually in hot water in the past decade. Similarly, Dr Frederick W. Seitz's masterly passivity, well suited to the booming budgets of the mid-sixties (but wearing thin towards the end) would have led him into trouble in the turbulent seventies. Handler's spell at the academy has been a much rougher ride. During this period, budgets have been falling, Congress has assumed much more explicitly than before the role of public watchdog over — some would say public prosecutor of — the scientific enterprise, which has also been threatened by the environmental nonsenses of the early 1970s and the fuss about genetic manipulation at the end of the decade. Inevitably, Handler has frequently been forced to play defending counsel. To his credit, he has done this job courageously — and in the process has made some stirring speeches.

Unlike the scientific academies of Western Europe, the United States academy has the seeds of dissension built into its constitution. Thus it resembles the academies of Eastern Europe in having a formal relationship with Congress, which is empowered to commission studies on whatever questions happen to flit across the scene. If, for example, Congress wishes to know whether small doses of radiation are dangerous, whether the accumulation of chlorofluorohydrocarbons in the atmosphere will dangerously deplete the ozone layer or whether the Department of Energy is telling the truth about energy forecasts, many committees of Congress will think first of commissioning a study from the academy. Technically, the work is done by the National Research Council, in whose management both the academies of engineering and medicine have a part, but in the public mind (and that of Congress) the academy of sciences is responsible. Since by definition the subjects on which Congress asks for advice are contentious, the academy frequently finds itself at the centre of public argument. Last year's report on the effects of small doses of radiation (see *Nature* **286**, 550; 1980) was a good illustration of the troubles that can arise. Occasionally, the academy asks for trouble, as with its own committee's pungent advice that people should not be unduly concerned with the exact composition of what they eat (see *Nature* **285**, 428; 1980). One of Handler's tasks, in the past twelve years, has been to give this stream of documents a collective stamp of authority. For the future, it would be better if the academy could arrange to shuffle off commissions for studies which are either pointless or untimely. Too much of what it has attempted should have been

refused, and should be in the future.

That is one task for Dr Frank Press. Another is to redefine the ground on which the scientific enterprise is defended. Hitherto, the academy's machinery for cogitating about problems of general concern has been directed towards specific contemporary problems, where data gathered by research promise to illuminate matters of general interest. The academy has been much less concerned with the machinery by which the scientific enterprise is sustained. The education of students in schools and colleges? The National Science Foundation has been allowed to set the pace, at least until its budget was slashed last month. The problem of university scientists, with or without tenure, engaged in the increasingly hazardous gamble of the pursuit of scholarship? Leave that to the American Council on Higher Education, which has strong opinions on issues such as tenure — and hope that people will one day recover from the over-abundance of documents churned out by the Kerr commission on higher education. The propriety of the arrangements academics make with their universities to share in the profits of extramural commercial enterprises? Or that universities make with commercial companies or with agencies of the federal government? The universities, the argument seems to go, are too jealous of their independence to welcome advice.

As events have turned out, the academy in Washington should move quickly to take up these and a host of related questions. Although the competitive universities and colleges of the United States are rightly proud of their autonomy, they are now afflicted by so many common and pressing problems that the academy's influence could be decisive. Should the protest against the new strict interpretation of the rules prohibiting the export of military material, which threatens to restrict the education of overseas graduate students, be left to a handful of college presidents (see *Nature* 9 April)? Are the new rules introduced by the Office of Manpower and Budget for accounting for the use of federal funds equitable or merely iniquitous? And will they achieve their declared purpose of saving federal dollars? Is it sensible that sources of financial support for graduate students should be as much curtailed as in the past few years? What — in the national interest as well as that of research — should be the balance in the funding of research between project grants (with their accompanying variable overhead percentages) and straight institutional support (of which there is at present almost none)? And if institutional support comes back into fashion, to which institutions should it be channelled by which agencies? Many of these are tedious questions, but they are much more important than the consequences of aerosols in the atmosphere.

The surprise so far is that the academy seems to have had very little interest in these issues. They are of course intangible and often boring. There is also always the danger that if a body such as the National Academy of Sciences speaks out forthrightly on these subjects, it will be accused of self-serving propaganda on behalf of its members. That risk is worth running. Indeed, the risk must be run. For with so many other well-wishers of the scientific enterprise emasculated by President Reagan's first budget, the academy is almost the only institution in Washington able to speak out. Moreover, the issues are likely to induce a harmony of purpose among the members of the academy sufficiently strong to carry the institution through the divisive issues certain to crop up in the next six years as in the past twelve — whether, for example, to protest at Sakharov's exile and, if so, how? Is this an agenda for Dr Frank Press?



Will the London Zoo survive (and how)?

The London Zoo, more formally known as the Zoological Society of London, is in serious trouble. In the annual report for 1980, published last week, the president of the society, Lord Zuckerman, says that the organization is "up against a brick wall". The evidence of the report leaves little doubt that he is correct. Last year, the society made a loss of £666,000, more than 13 per cent of its income of just under £5 million. This, however, is not the end of the tale. The society has made large losses in previous years — £130,000 in 1979, for example. The novel but ominous implication of last year's result is that the society's bank balance has been almost emptied, and replaced by a substantial overdraft. And that there are unmistakable signs that the chief source of revenue, the entrance fees paid by visitors to the two zoos, in Regent's Park, London, and at Whipsnade, is no longer an inexhaustible pot of gold. For at least two decades, as charges have been increased, the numbers of visitors to the zoos have been falling. Last year, attendance at Regent's Park fell by 11 per cent. Like many other enterprises in inflation-ridden Britain, the society (which is a charity and not technically a business) is being crippled by continually rising costs and the shortage of cash in the economy to spend on the services it provides.

The society's recipe for survival, outlined in the report for 1980, is logical enough but almost certainly destined for failure. The zoo, the argument goes, is constituted as a charity which provides public services of several different kinds — pleasure for visitors to the two zoos, educational experience for a great many school students and a substantial amount of scientific research, much of it distinguished and not all of it inspired by the veterinary needs of the animals on display. Last year, the two zoos managed to collect some £60,000 on account of their overtly educational work, by almost any test a miserly recompense for their educational value not merely in London but in the United Kingdom as a whole. Last year, the cost of research (£780,000) exceeded by £322,000 research grant income. So the society is asking why its value as an educational institution should not be more tangibly recognized by local and central government, and why a substantial part of the cost of its research must continue to be met by those who innocently spend an afternoon looking at exotic animals. On both counts, the zoo has a powerful case.

If the zoo were to go out of business — and if its financial position does not improve, it may have no choice — there would be a great outcry from all kinds of interested parties. Local education authorities would be prominent among those protesting that a great institution should not be allowed to sink without trace. They would be complaining that children in their schools would be deprived of an essential part of their education. The spendthrift among them would no doubt argue that if the London Zoo should be closed, it would be necessary for public funds to pay the cost of carting children to the distant game parks of central Africa. Some might even be able to raise the funds. The zoo's case is that if the local authorities are as certain as seems likely to mourn the disappearance of the London Zoo (and perhaps also of Whipsnade), they should contribute more generously now to help keep both institutions in being. The logic is impeccable. The snag is that local authorities are just now as hard-pressed as the zoo itself. And in the nature of British local government, they are incapable of making discretionary expenditure, however prudent. Voluntarily to shoulder part of the cost of supporting the London zoos would be politically risky, but the same local governments (the Greater London Council and the Inner London Education Authority) would have no qualms in foisting on their ratepayers the much larger cost of salvaging the society's collapse.

The obstacles to the zoo's case for more outside support for its research programme are a little less ominous. The point has already been established that for two years there should be a modest subvention (£134,000 last year) from central government funds, while the zoo's laboratories can compete for project grants

from public and private sources. The accounts for 1980 (which sensibly charge the cost of research "of direct benefit to animal management" against the income of the public exhibits) nevertheless show that the society had to spend £250,000 of its inadequate income on research. The principle that a private organization should qualify for a public subvention towards the overhead cost of research is unusual but is not an outrage. If a research programme is worthwhile, there is no reason why it should not qualify for public support as if it were carried out in some publicly owned laboratory. The obvious snag, that private management may mean fickle management, is easily overcome by making sure that all grants are short-term grants. In the zoo's present plight, the Advisory Board on the Research Councils, which unimaginatively cuts the annual research cake in Britain, should be willing to do more to help.

Public support of this kind, however, can only be a stop-gap. The zoo itself must acknowledge that its research institutes cannot prosper if they have to live from hand to mouth, keeping watch on the dwindling queues at the visitors' admission gates. It is now clear that the zoo's laboratories have a modest if distinctive contribution to make to some intriguing problems in animal physiology and veterinary medicine, but that they will never be comprehensive or dominating. In the interests of the laboratories, the zoo should therefore seek a more durable framework for its research. For a year now, there has been talk of a consortium including the zoo, the British Museum (Natural History) — which should simplify its name as well as root out cladistic excesses from its public exhibitions — and the Royal Botanic Gardens at Kew. All three institutions have in common their dual concern for public exhibition and scientific research. The connections between the three research programmes are by themselves, however, probably too tenuous to be a durable bridge. Nor is it likely, in present circumstances, that other British zoos would be willing to regard the London Zoo as a central research service for them all, writing cheques accordingly. The zoo's research enterprise might be better able to make its way in the world by setting itself up as a common service for academic and public research laboratories, for example by setting up well-run breeding colonies of primates. Either (or some other) way, the zoo must quickly work out a plan and set out to make it work. In the process, it may find it prudent to put more distance administratively between the research laboratories and the Zoological Society of London as a whole. There is no evidence that the fellows of the society who go to watch the pandas on Sunday mornings presume to interfere with the research, but these days a show of independence would do no harm.

Much the same would be necessary if the zoo were to succeed in recruiting support from local government to keep its zoological display in good shape. Local government politicians are more convinced than others that money means power, in this case influence over exhibition policy. Lord Zuckerman seems bent on persuading London's local politicians that they should follow the examples of their peers in Berlin, Rotterdam and elsewhere. Would he and the council of the zoo be prepared to share their present absolute power with those willing to keep the zoo out of the red? This unfamiliar thought is neither unthinkable nor unseemly. In the long run, it would help and not hinder if local governments were prepared to share anxieties about the zoo as well as support for it. Experience elsewhere suggests that the propriety of the display need not be damaged. The implication is that the zoo should evolve into a consortium of three separate charities — a research organization, a scientific society of some importance and a public exhibition. In these hard times, it is too dangerous to let the enterprise stand or fall as a whole. That way, it would be better organized for survival, better able to provide its well-wishers with a sense of belonging to a constituency. After all, many of the society's friends think it wrong that animals should be kept in zoos.

Vienna institute hit by espionage charges

Claims force Soviet official to resign

Allegations of espionage for the Soviet Union have forced the resignation of Dr Arkady Belozarov as secretary of the Institute of Applied Systems Analysis at Laxenburg, near Vienna. This dramatic development has intensified anxiety at the institute about its long-term financial security, already threatened by the Reagan Administration.

Dr Belozarov was appointed to his post at the institute in 1971, and took up his job in December that year. An article in the British news magazine *Now!* on 10 April said that Dr Belozarov, a plasma physicist by training, was also an agent of the KGB who had been the Soviet contact with a Norwegian double agent during his spell in Vienna. A statement put out by the institute on 15 April said that Dr Belozarov had offered his resignation and that this had been accepted.

The responsibilities of the secretary of the institute are largely those of external relations with member organizations (mostly national academies), potential donors (such as foundations) and the press. As with other senior posts at the institute, the secretary is appointed jointly by the director and the governing council. Dr Belozarov's name was the only nomination before the council in 1979. The institute's policy on recruitment requires that there should be a national balance among senior members of the staff.

The allegation in the *Now!* article is that Belozarov took over from another Vienna-based Soviet agent responsibility for keeping contact with a Norwegian petroleum engineer known as Jansen. According to the magazine, "Jansen" began working as a double agent for the Norwegian security services six years ago, providing a succession of Soviet contacts in Oslo and then Vienna with innocuous and sometimes misleading information about petroleum operations in the North Sea and the China Sea.

Dr Belozarov is reported to have met "Jansen" on several occasions in Vienna, to have been incompetent in making arrangements for meetings and to have been reluctant to pay for information provided until its authenticity had been checked. He is reported to have given a full account of his career as a member of the KGB and in the process to have revealed the identity of a Soviet agent in Austria. His meeting with "Jansen" arranged for 4 April outside the Volks Opera in Vienna is said to have been frustrated by the expulsion of another alleged Soviet agent

from Norway.

In his letter of resignation to the institute, Dr Belozarov protests that the allegations "are without factual basis", but that there is no impartial forum in which he can demonstrate them to be false. His letter says that he offered his resignation "in the interests of the institute".

The institute says that Dr Belozarov's conduct of his post as secretary had been "completely satisfactory" but that it has no way of independently determining the truth of the allegations against him and therefore had no choice but to accept his resignation. A spokesman said last week that in the present precarious financial circumstances, any member of the staff against whom allegations of this kind were made would be asked to leave.

One curious feature of the Belozarov affair, as reported, is that Soviet agents should have been willing to pay for information about oil technology in deep waters

such as those of the North Sea when the institute itself had only recently published an up to date account of the field.

The Austrian government, an enthusiastic benefactor of the institute, has reacted protectively. The Minister of the Interior, Mr I. Lanc, said last week that Belozarov had committed no offence under Austrian law and complained that "certain countries" not having common frontiers with the Soviet Union had come to undervalue detente and to regard the Institute of Applied Systems Analysis as an expendable target.

The institute's financial prospects are not nearly as bright as they may have seemed only a few months ago. Although the member organizations of the institute are national academies, most of these are reimbursed by their governments for the cost of membership. The Soviet Academy of Sciences and the United States National Academy of Sciences are the largest con-

Schuster quits as EEC research director

Brussels

The reshuffle of top jobs at the European Commission following the appointment of commissioners and the allocation of their portfolios at the beginning of 1981 has led to a flurry of resignations during the past two weeks, and rumours that some directorates-general are to be reorganized are gathering force.

Five directors-general are known to be leaving, among them Günter Schuster, head of Research, Science and Education (DG XII). His resignation was handed in amid speculation that he is dissatisfied with the way responsibility for DG XII has been split among three commissioners. At the beginning of the year, infighting among the new 14-man Commission caused a messy distribution of the research portfolios. Etienne Davignon (France) grabbed research and science, leaving Ivor Richards (United Kingdom) with education and training, and Karl-Heinz Narjes (West Germany) with industrial innovation.

Some members of the Commission claim that this arrangement is not as awkward as it might seem — decisions taken by the commissioners have in any case to be unanimous. The real reason for Schuster's departure may be that at 62 he would be due for retirement mid-way through the next four-year indirect action programme now being planned. A new man would be able to see the plan to completion.

His departure leaves a vacuum hard to fill. According to a close colleague, Herbert Allgeier, Schuster's greatest personal achievement was to build up from scratch the European Community's research policies from the shambles that followed the collapse of the Euratom programme at the end of the 1960s. Pressure is now being brought to bear on Davignon to

find a replacement as soon as possible, so that the new appointee can help draw up the new research policy guidelines.

Among the names being mooted is that of Fabrizio Caccia Dominioni, at present looking after research, development and nuclear policy. However, somebody may well be brought in from outside, although probably not another German or Briton. Leonard Williams of the



Schuster leaves the scene

energy directorate-general is being replaced by Christopher Audlands. In the best Byzantine traditions, the top jobs are distributed so that national pride and influence are not diminished with each four year reshuffle.

Other changes are rumoured, but one decision, it seems, has definitely been taken. The Environment and Consumer Protection Service (ECPS) is to be elevated to directorate-general status thanks partly to the patronage of the European Parliament. The ghostly DG XI, which was left vacant after an earlier round of administrative musical chairs, will once more have an occupant. Michel Carpentier will remain in charge as head of ECPS and will henceforth be officially addressed as director-general.

Jasper Becker

tributors at rather more than £1 million each a year.

In the United States, the National Academy's annual contribution has since the inception of the institute been met by the National Science Foundation from its allocation of funds for international activities, cut in President Reagan's first budget from \$15 million to \$10 million for the coming financial year. After protests from the academy that the statutes of the institute require twelve months' notice of withdrawal, the putative contribution for 1982 has now been restored to the foundation's budget, but on the strict understanding that this will be the last contribution to be sanctioned by the Administration.

In principle, there is no reason why the academy should not seek to raise its annual contribution from some other source than the United States government, but in present circumstances this will not be easy. The 1982 budget for the institute will be agreed between the member institutions at two council meetings to be held in June and November this year. The United States academy will have to give notice of intended withdrawal at the end of this year if it is not by then assured of funds with which to pay its 1983 subscription.

It is ironical that the institute, founded partly with the intention of providing a place at which people from East and West could work on problems of international importance, should have run into trouble because of allegations of espionage.

US graduate employment

Industry has jobs

Washington

Despite continued uncertainty in the US economy, business is booming for engineering graduates — and natural scientists are not far behind. Figures just released by the College Placement Council of Bethlehem, Pennsylvania, reveal that engineers who expect to graduate this summer are already being offered starting salaries of up to \$26,000, a 10 per cent increase over last July. Graduates in physical and Earth sciences were being offered \$22,000, an 18 per cent rise compared with last year.

This rapacious demand for technical graduates has been confirmed by a report from a New York consulting firm, Deutsch, Shea and Evans, which claims that there has been a sharp up-swing in employment prospects for graduate engineers since last November. And the figures are in marked contrast with starting salaries for arts graduates, which average about \$12,000, with many failing to find any jobs at all.

The shortage of engineers is partly a reflection of declining enrolments in university engineering schools. For example, 2,751 engineering doctoral degrees were awarded in 1980, compared

with 3,074 in 1972; although this also reflects the fact that fewer engineers are staying on to do graduate work, the 1980 figures include more than 60 per cent foreign students, many of whom were not expected to stay in the United States.

Recruitment difficulties have also produced near-crisis in the engineering schools, few of which can afford to match salaries in industry. According to William A. Cox, president of the National Society of Professional Engineers, there are now 1,800 unfilled faculty positions in engineering schools — a situation which has led the society to add its voice to those asking Congress for federal assistance to upgrade facilities and research equipment to make the educational environment more attractive.

There is also a feeling that the turnaround in demand for technical graduates reflects greater optimism in the economy than six months ago. "In general, the market in technical areas is the best that it has been for years", says Victor Lindquist, director of the placement office at Northwestern University which carries out regular surveys of employment prospects. He claims, for example, that metal companies such as mining and steel, sectors which are likely to benefit directly from a significant increase in military spending, seem to be stockpiling talent in anticipation of a quick recovery in the economy.

The College Placement Council survey revealed that, as in the past, engineers represent the hottest buy in the employment market. More than 63 per cent of the job offers reported in the survey as having been received by mid-March were to future engineering graduates, although these only made up 7 per cent of the bachelor degree graduates in 1980–81.

In proportional numbers of offers, the scientific disciplines were dominated by computer science graduates, another field where demand is considerably in excess of supply, a point which has been made recently by the National Science Foundation. However, the average starting salary offered to such graduates of \$19,968 was considerably lower than the \$21,912 offered to graduates in physical sciences.

Among masters level graduates, the highest offers went to chemical engineering graduates, with an average starting salary of \$26,240 — almost 13 per cent higher than the figures last July. Geology and related geological sciences, in strong demand in the petrochemical and mining industries, ranked second with an average salary of \$25,380. All these figures are based on offers reported by placement offices at 161 US colleges and universities.

Figures released last week by the National Science Foundation reveal that there is an increasing tendency for graduate scientists, engineers and technicians to enter service rather than manufacturing industries. These industries accounted for 60 per cent of the total growth of employment for graduates in the three categories

between 1978 and 1980, even though only accounting for 40 per cent of the total employment.

Total science and engineering employment in private industry was virtually unchanged between 1970 and 1980, but this trend may be reversed, however, if the current patterns of employment for this year's graduates is maintained.

David Dickson

Soviet space programme

No shuttle yet?

The high cost of space research is troubling Soviet planners as well as their counterparts in the space agencies of Europe and the United States. At a ceremony on 10 April commemorating the twentieth anniversary of Yuri Gagarin's first manned space flight, the president of the Soviet Academy of Sciences, Dr Anatoli P. Aleksandrov, spoke in a low key for such an occasion.

Aleksandrov's speech was little more than a justification of expenditure so far, and a defence of the proposition that the more that is spent on space, the greater the eventual benefit. This is not the first indication that Soviet space research may be running into financial difficulties. Earlier this year, the Soviet side had to make it clear to their French partners that if the two countries were to carry out a joint mission to Halley's comet, spending on their Venus mission of 1986 would have to be halved.

To demonstrate the priority being given to cost-effectiveness, Aleksandrov explained that three and a half years ago the cosmonauts aboard Salyut-4 were working on behalf of only 40 scientific and industrial bodies, whereas in the latest mission, Vladimir Kovalenko and Viktor Savinykh on Salyut-6 were providing data for some 400 organizations.

Naturally Aleksandrov did not discuss any military aspects of the Soviet space programme, although he did add to the abuse of the US space shuttle as a military vehicle. The press in the Soviet Union suggest that the US shuttle is primarily for launching laser weapons into space.

Soviet intentions on reusable launching systems remain unclear, however. Aleksandrov did not discuss any such plans, but Soviet cosmonaut Vitalii Sevastyanov stated in the Slovak daily *Pravda* that the Soviets were considering the construction of a shuttle. He claimed, though, that the present conventional methods used in Soviet space launches were more convenient than a shuttle for their programme based on large orbital stations designed for "peaceful purposes" and "suitable for a longer stay in orbit of scores of specialists".

There is speculation in the West, however, that at least one experimental Soviet shuttle has been launched — number 929 in the "Kosmos" unmanned satellite series.

Vera Rich

Anglo-German project

Aurora held up

Red tape has already delayed the British half of an Anglo-German programme of auroral investigations for several months. Last week, one of the obstacles — agreement on the wording of a lease for land at a site in the north of Scotland — was surmounted. But a further six months delay is now likely, pending the installation of an electricity supply to the site.

The joint programme, known as SABRE, is for radar backscatter observations of the aurora from two sites — near Wick in Scotland and at Uppsala in Sweden. The two collaborating groups are led by Professor Tudor Jones at the University of Leicester and Professor I. Axford at the Max Planck Institut für Aeronomie, Lindau. The British Science Research Council and the Max Planck Institut are supporting the experiments to the tune of about £200,000 each.

The programme is designed to complement a similar twin auroral radar, called STARE, operated from two sites in northern Scandinavia by the Lindau group. Observations from the Swedish site began last January, but delays in preparing the Scottish site have meant that spatially and temporally resolved observations have not yet been possible.

The Scottish site is one of four sites originally owned by the Ministry of Defence but now controlled by the Property Services Agency which deals with matters relating to public land. Most of the hold-up seems attributable to delays in reaching agreement on improving accommodation and connecting an electricity supply. The agency, which plans to sell the other three sites, has been concerned that the electricity supplies of each should be independent. At present, the supply to the site which the University of Leicester wants to lease is routed via a transformer situated elsewhere. Four options for installing a new supply, presented by the North of Scotland Hydroelectric Board, have been turned down as too costly.

Some blame the delays on the snail's pace at which the Property Services Agency works. After 18 months of turning the bureaucratic handle, an acceptable draft lease finally arrived at the University of Leicester only last week. But the problem of installing a new electricity supply remains. A mains cable supplying the site was apparently damaged a few years ago when a new road was built. Repairs involve laying new cables. The North of Scotland Hydroelectric Board estimates that the job could take six months and cost about £5,000. The agency has now handed over responsibility to the university.

If work cannot be speeded up, a temporary generator will have to be installed if experiments are to begin before winter.

Judy Redfearn

US nuclear power

Cost disputed*Washington*

Efforts to meet the safety problems created by nuclear power have already made nuclear energy considerably more expensive than electricity from modern coal-fired plants, even with the sophisticated anti-pollution equipment needed in coal-fired stations. And the gap is likely to increase at least until the end of the decade according to a report published in New York last week by an independent energy economist, Charles Komanoff.

The nuclear industry has reacted sceptically to Komanoff's prediction, claiming that he underestimates the productivity of nuclear power, and overestimates the productivity of coal generation. The Atomic Industry Forum, chief spokesman for the nuclear industry, issued a statement saying that the conclusions clash with those of other economists, and that "to predict cost trends for coal and nuclear in a period of fuel-price volatility and still-uncontrolled inflation seems highly dubious."

However, Komanoff, who last year gave evidence on nuclear power costs to the UK House of Commons Select Committee on Energy, is standing by his figures. He points out that similar scepticism greeted a prediction he made five years ago that the nuclear industry would be operating at only 55 per cent of capacity through to the end of the 1970s, a figure which has turned out close to the truth.

He also claims that the argument about coal being cheaper than nuclear power appears to be supported by figures published by the Atomic Industry Forum itself. In a report issued two months ago, the industry group estimated that electricity from the three nuclear plants which started operation in 1978 cost 2.5 cents per kilowatt hour to produce, compared with 2.0 cents for coal and 5.7 cents for oil.

Komanoff maintains that the more nuclear power has expanded, the more generic problems it has revealed, and the greater the effort required to prevent the number of nuclear incidents from increasing at the same rate. He says this is substantiated by the steady increase in the number of warnings of potential generic problems by the Nuclear Regulatory Commission during the past decade.

Komanoff bases his future projections on the cost of obtaining electricity from coal-fired plants by assuming that increasingly stringent controls will require the use of scrubbers to remove sulphur dioxide — devices that can consume up to 25 per cent of the capital costs of a new plant — and that the cost of coal will increase at a rate about 2 per cent more than the inflation rate. On this basis a coal plant will increase in capital cost from \$583 to \$794 per kilowatt hour of generating capacity in "constant" dollars. In

contrast, extrapolating on past trends in the rising costs of nuclear power plants due to the need to include increasingly stringent safety precautions, Komanoff predicts that the cost will rise from \$887 to \$1,374 per kilowatt hour of capacity.

The Atomic Industry Forum is looking closely at Komanoff's figures. Its own analysis, based on power plants coming into operation between 1970 and 1978, is that "nuclear electricity in 1979 continued to show a modest advantage in total generating cost compared with neighbouring coal units", and that nuclear power also compared favourably with coal in capacity factor. At the same time, the industry group admits that regulatory delays and public intervention in licensing procedures "run up nuclear costs significantly"; it claims that there are approximately one dozen nuclear plants currently snarled up in bureaucratic delays which will cost US consumers as much as \$3,000 million in higher utility bills.

Whichever analysis of the relative costs of nuclear and coal is correct, however, both sides agree that immediate prospects for the US nuclear industry are bleak. Only 13 new nuclear-power plants have been ordered in the United States since 1975, and plans for 50 plants have been cancelled — many because of decreasing predictions of future energy needs.

On top of this, no new plants have been ordered in the past two years. The nuclear industry is hoping that President Reagan will make good his election promises of increased support, and is already providing him with suggestions as to how this might be achieved, from revamping some of the safety regulations which Komanoff claims to be at the root of the cost escalation, to proposals for federal financial aid for nuclear-plant construction.

Overhauling the regulatory apparatus, however, will not be easy. Only one of the five seats on the Nuclear Regulatory Commission is vacant, and two of the existing commissioners have been harsh critics of some of the industry's demands. In addition, Mr Reagan's rumoured choice for the chairmanship of the commission, former head of naval operations Admiral James L. Holloway, is reported to have turned down the offer, and with no clear indication of the direction in which it is meant to be moving, the commission is running at half speed.

David Dickson

Anti-viral drugs**Herpes on trial**

Volunteers, 50 male and 50 female, with genital herpes simplex infections are being sought by the Infectious Diseases Unit at the University of Texas Health Science Center at Dallas to take part in clinical trials of the promising antiviral drug acyclovir (acycloguanosine). There is as yet no effective treatment for genital herpes, which is sexually transmitted and now

considered to be of epidemic proportions in the United States, where 300,000 new cases are reported each year. Each new patient becomes a carrier for life. Acyclovir has generated great enthusiasm at Burroughs Wellcome, its developers, as a real advance in treating the disease. One worry, though, is that viruses may acquire resistance.

Herpes simplex virus (HSV) can infect any part of the skin or mucous membrane. HSV Type 1 cause the familiar "cold sore" blisters on the lip. Genital herpes, caused by HSV Type 2, is a more severe condition and causes painful blisters in the genital area. In the Dallas trial the drug will be applied as an ointment over skin areas where patients feel a tingling, burning or itching sensation (a "prodrome") suggesting that a blister is about to develop. Acyclovir will be compared with placebo cream containing no active agent. If the drug succeeds in reducing the period of recurrences to one or two days, it should not only reduce the period of pain for the patient, but also reduce the capacity for passing on the disease to others.

Because viruses are parasites of the normal systems of the cell they infect, antiviral drugs tend also to act against the host cell. There are few points at which the life processes of the virus can be properly differentiated from those of the host, but acyclovir acts specifically against herpesvirus through one such reaction (*Nature* 272, 583; 1978). It is phosphorylated by the herpes-specified enzyme thymidine kinase, so that the active triphosphate derivative is present in much greater quantities in infected than in uninfected cells. In a sense, the virus thereby invites its own destruction. It remains to be seen whether acyclovir will prove active against even the dormant form of the virus. Herpes infection is characterized by the way in which all outward evidence of the infection can disappear, but the virus persists undetected in certain nerve cells.

Acyclovir is being used as an ointment in the Dallas study, but later tests will include systemic (for example intravenous) injections which could increase the range of infections against which the drug can be used. Already acyclovir has proved active as an eye ointment against herpes simplex corneal ulcers, and less conclusively against viral pneumonia.

The US National Institutes of Health are to fund trials of intravenous acyclovir in newborn babies with neonatal herpes, in comparison with adenosine arabinoside (vidarabine), and in cases of herpes encephalitis, in comparison with the adenoside of adenosine monophosphate. Neonatal herpes is a herpes simplex infection contracted from the infected mother during passage through the birth canal — if the virus reaches the brain it can cause death. Herpes encephalitis is caused by reactivation of the virus in the brain.

Charles Wenz

European Earth sciences Unity prevails

Strasbourg

Europe's Earth scientists may yet unite. A meeting on 13–16 April at the assembly of the Council of Europe saw the first public appearance of the European Union of Geosciences (EUG). At present, EUG consists of a founding committee and has the official status of a working party of the Parliamentary Assembly of the Council of Europe. Its aim is to provide an organization which encourages the exchange of ideas and information across the more traditional subject barriers that separate geologists from the geochemists, geophysicists and oceanographers. To that end, membership is open to any geoscientist, but officials of EUG are to be nationals of the member countries of the Council of Europe.

At present, EUG is run by a founding committee under Professor C. J. Allègre of the University of Paris, and its formal statutes are to remain provisional until approved at a future meeting. EUG intends to hold a regular meeting in the spring in Strasbourg every other year; its interdisciplinary scope is to be complemented by more specific topical conferences to be held between the more general meetings. The concept and format of EUG owe a lot to the success of the American Geophysical Union and its big annual spring and fall meetings, which have almost become a must for many European researchers. However, the scope is to be wider insofar as it is to encompass all Earth scientists, not only those in the geophysical community. In this way, it hopes to avoid rivalries and misunderstandings with the European Geophysical Society, which caters for the same scientific community as the American Geophysical Union.

Although the European Science Foundation (ESF) did not play a major part in the organization of the Strasbourg meeting, it appears that ESF is prepared to help set up and support a small permanent office for EUG in Strasbourg. Is this the beginning of ESF taking a more active interest in the Earth sciences, an area hitherto neglected by its working parties?

The European Parliament building was made available by the Council of Europe for this first EUG meeting. Although organized at relatively short notice, it nevertheless managed to attract more than 600 participants and enough presentations to run a full four-day programme grouped in 16 different sessions.

The abstracts from the first meeting appeared as the first issue of an ambitious new publication called *Terra Cognita*. EUG hopes to develop this into a quarterly journal which should serve as a forum for scientific information on research projects and activities as well as to promote the discussion and exchange of ideas within the Earthsciences community. Konrad Guettler

Out on a limb

Brussels

Dutch Socialist Member of the European Parliament, Annie Krouwel-Vlam, is trying to whip up support for a European Community directive on common rules for organ transplants. She is surprised that there are no rules regulating relations between organ banks and facilitating access for would-be recipients of new parts. Such measures would implant a greater awareness of the benefits of the European Community in the breasts of its citizens, she asserted.

Her suggestion has been passed to the Community's organ banks and the European Commission. So international banking could soon take on a new meaning.

Jasper Becker

Interferon business

Trials ahead

Rotterdam

That interferon is big business can only have escaped the notice of those totally occupied for the past two years in preventing the pollution of space by small plastic tiles. That it continues to be big business was evident from the number of companies represented last week (19–24 April) at the interferon meeting organized by TNO, the Netherlands organization for applied scientific research.

Half of the British participants were from Beechams, Celltech, Fisons, ICI, G.D. Searle and Wellcome, with about the same proportion of commercial people among the West Germans. The large United States contingent included representatives of Cetus, Du Pont, Enzo Biochem, Electro Nucleonics, Genentech, Hoffman LaRoche, Hydron, Meloy, Sepetex, Schering and Shell. Not a slide went unphotographed nor a talk unrecorded by one or more commercially interested party.

Many of these companies already had a declared interest in interferon but some were new to the game. ICI let it be known that it has cloned an interferon gene from lymphoblastoid cells in collaboration with Dr E. de Maeyer's group at the Institut Curie in Orsay. ICI says that this is just an exercise in cloning, but with the setback in the fortunes of the company earlier in the year (*Nature* 12 March), there must be thoughts of trying to cash in on clones.

There is little doubt that the demand for highly purified interferon is growing, but the growth can be fitful. Wellcome, for example, recently claimed that supply had outstripped demand for its alpha-interferon (from white blood cells). That situation could easily be reversed if the current clinical trials of Wellcome's interferon show promise. The interim results, reported in Rotterdam, of several relatively large trials of a variety of different alpha-

interferons suggest that it has a tantalizing degree of activity against a broad spectrum of viral diseases and tumours. But activity is not enough, or consistent enough, in most diseases to give a final verdict without more trials. Also, as genetically-manipulated bacteria start to yield sufficient amounts of the ten or so subtypes of alpha-interferon, the permutations of subtypes and diseases becomes huge.

Meanwhile, large-scale trials of beta-interferon (from fibroblasts) lag slightly behind. The largest has yet to start. Financed and organized by the National Institutes of Health (NIH), the trial will use 50,000 million units of beta-interferon, enough for several hundred patients. The interferon is being supplied by Flow Laboratories Inc., which won the \$2 million contract a year ago. First deliveries were due a few weeks ago, but NIH appear to be in no hurry. This is lucky for Flow Laboratories, whose own production has not been going smoothly. Fortunately, they have a subcontract with the West Germany company of Dr Rentschler Arzneimittel GmbH which has continued to amass beta-interferon made by its own, less advanced, techniques.

The next big contract, also for NIH, will be for gamma-interferon (from cells of the immune system). Far less is known, at least publicly, about this type of interferon, but Biogen and Genentech are rumoured to be on the verge of cloning the relevant genes.

Other commercial interests centre on monoclonal antibodies against the interferons. Several new monoclonals against either alpha or beta-interferons were announced in Rotterdam, but only the original monoclonal (see *Nature* 285, 446; 1980) is as yet commercially available. The sellers are the British company Celltech Limited, so far its only product. It will be surprising if the Celltech catalogue does not double soon.

Peter Newmark

Dioxin hazards

Secrecy at Coalite

Secrecy about the medical consequences to workers exposed to dioxin continues at the British company Coalite and Chemical Products Limited. Documents now to hand show that Coalite will not agree to the further medical investigation of workpeople if the physicians concerned insist that the results are eventually published.

Coalite's shyness is not unprecedented. Last year, the company withheld information on an epidemiological survey of workers exposed to dioxin (2,3,7,8-tetrachlorodibenzodioxin) at its Bolsover (Derbyshire) plant in 1968-71 (see *Nature*, 6 March 1980). The exposure followed an explosion in a vessel used for the manufacture of 2,4,5-trichlorophenol and the workers affected by dioxin developed the skin disease chloracne.

The fresh trouble which has arisen at Coalite results from the need for more

detailed investigations of individual workers to make up for the deficiencies of the epidemiological survey of 1977-78. That involved only 41 of the 90 workers known to have been exposed, but revealed (almost a decade after the event) some abnormalities of blood chemistry and immune function — but none of the expected chromosomal abnormalities in circulating lymphocytes.

The news that Coalite was unwilling to disclose information about this survey served to alert the statutory Employment Medical Advisory Service of the Health and Safety Executive to the existence of such data, which was disclosed on a confidential basis after three weeks. The service has also persuaded Coalite that its medical adviser, Dr George May, should publish some of the results.

The employment service has in the meantime been uncomplimentary about the survey which, it says, is deficient in its failure to follow up past employees and in the inadequate matching of the control populations. The survey covered 41 workers known to have been exposed to dioxin, 54 who might have been exposed and 31 intended as a control group but drawn for convenience from senior office, laboratory and works staff, most of whom were being investigated for blood lipids at the time.

The 1977-78 survey showed blood-lipid disorders, with elevated concentrations of serum cholesterol and triglycerides, and depressed levels of high-density lipoprotein. The enzyme γ -glutamyl transpeptidase, diagnostic of liver function, was abnormally high in the exposed group. On the other hand, levels of two immunoglobulins (IgD and IgM) were depressed in the exposed group, suggesting impairment of immune as well as liver function.

The significance of these findings is far from clear, with the consensus among physicians in the field that further investigations are needed. They argue that the persistence of the phenomena uncovered by the earlier study should be confirmed — but also that further surveys should be more carefully controlled.

Dr George May, Coalite's company doctor, confirms that those workers included in the original survey have since been offered a second medical examination, but that "not everybody availed themselves of the opportunity". In the event, 29 of the original sample of 41 workers were re-examined, but measurements of immunoglobulin were not made.

There are no plans, however, for further investigations of the Coalite workers, according to Dr George Sorrie, deputy director of the Employment Medical Advisory Service, who says that his service has not been presented with any "realistic and useful" proposals for a long-term assessment of the health of the Coalite workers. The reassessment of the state of health of dioxin-exposed workers at

Dioxin a sarcoma risk

Reassessing clinical data is not the only reason for monitoring the health of workers exposed to dioxin. Recent letters in *The Lancet* (P. A. Honchor and W. E. Halperin, i, 268, 31 January 1980; and R. R. Cook, i, 618, 14 March 1980) highlight another reason.

The authors of the letters point out that in four industry-sponsored mortality studies of dioxin-exposed workers in the United States, 105 deaths were recorded. In four of these cases, death was caused by soft tissue sarcomas. These are rare forms of cancer and for most the aetiologies are unknown.

Soft tissue sarcomas have also been identified in another cohort of workers considered to have been exposed to dioxin. According to two Swedish epidemiologists, Dr Lennart Hardell and Anita Sandstrom of the Centre of Oncology at the University of Umea, exposure to the dioxin contaminated herbicide 2,4,5-T (2,4,5-trichlorophenoxyacetic acid) or chlorinated phenols 6 to 20 years previously could be responsible for the sixfold greater incidence of soft tissue sarcomas recorded in a survey of Swedish forestry workers (*British Journal of Cancer* 39, 711; 1979).

Alastair Hay

Coalite is thus far from complete. One clinician describes the re-examination of the 29 workers as a "totally useless piece of work" which cannot be expected to yield any meaningful results.

Coalite's managing director, Mr Charles Needham, has been approached by several clinicians prepared to carry out further investigations. Coalite's reluctance to agree to publication of the results from any future comprehensive medical screen of its workers has, however, forced at least one clinician to abandon his proposal.

Mr Needham, described by his secretary as solely responsible for press relations, has not been available for comment. One consequence of criticisms of the 1977-78 survey, and of Coalite's continued insistence on secrecy, is that the financial sponsor of the survey (an independent non-governmental source) will be unwilling to provide further support.

As the law stands, Coalite may have the last word. Hitherto, the company has been able to exploit an ambiguity in the 1974 Health and Safety at Work Act when deciding whether or not to release medical information on its workforce (see *Nature* 6 March 1980, p.2). The routine powers of the Health and Safety Executive to demand medical records of workers likely to have been affected by a product lapse when manufacture stops. Although Section 27 of the act allows the executive to disclose whatever information it requires (see *Nature* 1 May 1980, p.4), epidemiologists at the executive consider that a test case is required to clarify the law.

Alastair Hay

CORRESPONDENCE

Majority verdict

SIR — In a recent communication to *Nature* G. G. Simpson writes about a small group of cladists at the American Museum of Natural History that is not representative of the staff as a whole (*Nature* 26 March, p.286). In his view a majority of the staff rejects cladistic or Hennigian procedures. There are 32 professional zoological systematists on the staff here, not including emeriti or honorary staff appointments, and 26 of these use cladistic methods in their systematic work. The small, unrepresentative group referred to by Simpson therefore amounts to 81 per cent of the relevant staff. On the recommendation of three scientific departments, the Council of the Scientific Staff, and the Awards Committee, the trustees of the museum in 1975 gave Willi Hennig its Gold Medal for distinguished contributions to science.

J.W. ATZ, C.J. COLE, N. ELDRIDGE,
W.K. EMERSON, E.S. GAFFNEY,
B.N. HAUGH, L.H. HERMAN, E. KIRSTEUEER,
M.C. McKENNA, J.G. MAISEY, C.W. MYERS,
G. NELSON, N.I. PLATNICK, F.H. RINDGE,
D.E. ROSEN, J.G. ROZEN, R.T. SCHUH,
C.L. SMITH, I. TATTERSALL, R.H. TEDFORD,
R. WYGODZINSKY AND, R.G. ZWEIFEL
*The American Museum of Natural History,
New York, USA*

"I was there"

SIR — Your account (*Nature* April 9, p.435) of a lecture I delivered "at the Rockefeller University in New York last week" raises in me the strong suspicion that your reporter did not hear what I said. This impression is strengthened by the fact that the lecture he reports was actually given in the Cornell University Medical Center. J. L. GOWANS
*Medical Research Council,
London W1, UK*

Dr Gowans' suspicion is unfounded. Our reporter indeed heard what was said and reported accurately that part of Dr Gowans' talk concerned with the applicability of the Rothchild principle to the financing of medical research in Britain. The Cornell University Medical Center is on the other side of York Avenue in Manhattan from the Rockefeller University, which had advertised the meeting at which Dr Gowans spoke. The two institutions run many joint seminar programmes, but are of course organisationally distinct — Editor, *Nature*.

Evolution's Waterloo

SIR — While no sensible person could disagree with your position on the "theory of evolution" (*Nature*, 12 March, p.75) I think your arguments are unnecessarily weakened by the sloppy use of words in multiple meanings, enabling your opponents to confuse the issue.

The phrase "theory of evolution" for example, is used in at least two quite different meanings. At one time it means "a record of events in the history of the Earth", at another it means "an explanation of the underlying causes of these events".

The term "metaphysical theory" is another example of this abuse. Popper notwithstanding, a metaphysical theory, *sensu stricto*, is a theory which cannot be verified (or

falsified) either now, or ever or anywhere, say the proposition "God exists". The record of historical events, say the proposition "Napoleon lost the battle of Waterloo" is not a metaphysical theory. Although the event cannot be rerun, it certainly can be verified in many different ways. Therefore the "theory of evolution" in its first meaning cannot be called metaphysical, technically or otherwise. An explanation of why Napoleon lost the battle of Waterloo, on the other hand, may or may not be metaphysical, according to whether it is based on verifiable facts or not. An explanation based on the intervention of an army of angels could be called metaphysical; one based on the superiority of the British and Prussian rifles may or may not be true, but it certainly is not metaphysical. The same reasoning holds for the "theory of evolution" in its second meaning.

*Administration de l'Industrie, S. V. VAECK
Brussels, Belgium*

SIR — As an American "cladist" and evolutionist concerned with the health of both disciplines and the public reaction to them, I feel that some discussion of the issues raised in *Nature* is warranted.

The issue of creation versus evolution entails two points, neither of which was mentioned in sufficient detail for the public readers of *Nature* properly to understand. First, the real controversy is over whether or not the phenomenon of evolution occurs. It does not concern whether Darwin or anyone else was correct about the mechanisms behind the phenomenon. To imply that a refutation of Darwinism or neo-Darwinism constitutes a refutation of evolution is like implying that a refutation of Newtonism is a refutation of gravitational attraction. Both conclusions would be misconceived. Second, the issue of creationism versus evolution certainly does not rest on whether Karl Popper's particular philosophy of science renders both creationist myths and Darwinian conjectures "untestable." There is a higher principle to resolve this issue.

Science is a discipline of open-minded inquiry which seeks to provide explanations about the world without invoking the supernatural. Creationism as practised by fundamentalist Christians is a close-minded doctrine built around an ultimate source of truth, the Bible. Creationism is dedicated to the proposition that the only explanations about "origins" are the supernatural explanations set forth in their ultimate source of truth. Thus "scientific" creationism, by any standard of rational scientific philosophy, is not science nor can it ever be science. Karl Popper is quite aware of this and has said so in *Conjectures and Refutations*.

Another issue concerns what is being called "cladistics". Scientists who fall under this label have diverse views about the proper relationship between "cladistics" and evolution. As one "out-and-out cladist" I cannot imagine something less interesting than a branching diagram (cladogram) without an evolutionary interpretation. Indeed, I cannot imagine beginning a study of biogeography without such an interpretation. I will even claim that an assumption of descent with modification is necessary to do such analyses. I base my claim on two points.

First, I have never seen a natural hierarchy

that was not due to historical descent in the larger sense of the phrase. This includes natural hierarchies in both the organic and inorganic worlds. Examples of natural hierarchies in the organic world are, of course, phylogenies of both organisms and languages, and the hierarchy of ontogeny of individual multicellular organisms. The only example of a natural hierarchy in the inorganic world that I know of is the hierarchy that continents display through descent from common "ancestral" continents. We observe the unfolding of ontogeny; we have detailed histories for the evolution of languages; we have a very good mechanism for the drifting of continents. Are we to think that the only other example of a natural hierarchy, the branching relationships between species, is different in kind from the other three?

Second, the only alternative to the evolutionary interpretation is the assumption of a logical creator who made each species (or group of species, or language, or continent) to look as if it evolved. As I wish to practise science rather than theological metaphysics, I make the assumption of evolution. If I wished to practise theological metaphysics, I would certainly not conceive of a logical creator who was, at the same time, tricky enough to make things look as if they evolved when they did not. I make the assumption of evolution without letting my hypotheses of evolutionary relationships (cladograms to my mind) be dictated by particular theories such as neo-Darwinism. If I let neo-Darwinism dictate *a priori* what relationships were possible then I would be no better than the fundamentalists who let the Bible dictate to them what is possible.

E. O. WILEY
*Museum of Natural History
and Department of Systematics & Ecology,
University of Kansas, USA*

No to sociobiology

SIR — The recent controversy on genes and racialism (*Nature* 22 January, 12, 19, 26 February) may have some relevance to caste system in India, since it has now assumed a great political significance and is creating social tensions. Although castes are manmade (all Hindus are born unequal!), and the system is nowhere near as old as the human race, yet attempts are often made to establish a biological basis of castism in line with the insect societies. Wilson (in *Sociobiology*) has refuted any genetical basis for caste system, yet willy-nilly, theoretical biology has become a convenient and manipulative tool to disdain or support such social practices.

The problem will become more complex and acute if attempts are continued to biologize every social attribute, be it racialism, castism or nazism. Biologization may be well-intentioned, and it may well prove that some of our beliefs are wrong and unscientific, yet by this process it may generate unnecessary, avoidable controversies and social conflicts. Surely, without biologizing, one can distinguish between races and racialism and between castes and castism. One may dispute or not give a damn if the human race is united or divided, but one can surely condemn racialism or castism without looking for biological evidence or defending the theory of kin selection. Let us keep sociology and biology apart.

B. BANERJEE
Tea Research Association, Assam, India

NEWS AND VIEWS

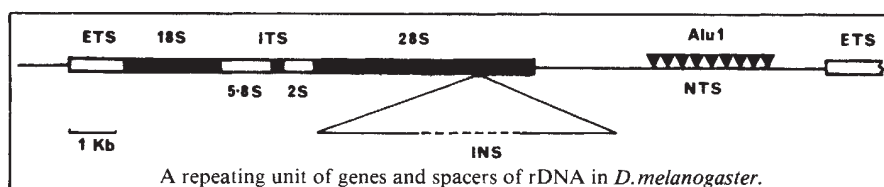
Springcleaning ribosomal DNA: a model for multigene evolution?

from Gabriel Dover and Enrico Coen

THE nuclear DNA of higher organisms is subject to such a diversity of sequence rearrangements that one might wonder how developmental or evolutionary stability is maintained. The variety of rearrangements, no doubt arising as errors in the regular processes of DNA replication and recombination, gives rise to seemingly randomly distributed families of multiple-copy sequences. A major unresolved question concerns the biological and evolutionary consequences of these mutational processes. One particular family of sequences that might provide the first insights into the rates at which different mutational mechanisms progress and of their possible phenotypic effects is composed of tandem head-to-tail arrays of ribosomal RNA genes (rDNA). From the data reported in this issue of Nature, p. 749, by H. Roija, J.R. Miller, L.C. Woods and D.M. Glover, and elsewhere²⁻⁵ it is now possible to observe in one family of repetitive sequences most, if not all, the phenomena usually depicted separately in different families.

The 18S, 5.8S and 28S RNA moieties of eukaryote ribosomes are co-transcribed from a unit of contiguous genes separated by internal and external transcribed spacers (ITS and ETS in the figure). Units are separated by non-transcribed spacers (NTS), in arrays that are located at the nucleolar organizers (NO). An exhaustive *variorum* account of the organization of rDNA and other multiple gene families has recently been compiled by Long and David⁶.

Since the pioneering studies of D.D. Brown, P.K. Wellauer, R.H. Reeder and I. Dawid into rDNA variation in species of *Xenopus*, it has become apparent that this family of repeats has been subjected to a new type of evolutionary force termed horizontal^{7,8} or concerted² evolution. This term is used to describe extensive sequence homology of NT spacers within a species



even though there is very little homology between spacers of related species. The mechanism responsible for concerted evolution of copies of a sequence within and between individuals should take into account the striking interspecific homology of the genes with which the spacers are regularly interspersed — for example^{3,9}, there is 93 per cent homology within 200 base pairs of the 28S gene of an insect³, an amphibian⁹ and a ciliate protozoan¹⁰ — and should result in rapid 'horizontal' spread of each new spacer variant to all rDNA units of all individuals of a species. Alternatively, it is conceivable, although unlikely, that both the rDNA arrays and populations go through drastic reductions in number followed by the establishment of new arrays from a few founder units. The recent analysis of sequence variation of rDNA in a variety of organisms, in particular in species of *Drosophila*, reveals several overlapping levels of evolutionary change that throw some light on the mechanism of concerted evolution and the complex progression of rDNA in general.

In *Drosophila melanogaster* there are two clusters of rDNA situated on the X and Y chromosomes. An interesting additional feature is that 60 per cent of the 28S genes of the X chromosome contain an insertion (INS-type 1). Another non-homologous type II INS is present in 15 per cent of the 28S genes of both chromosomes. Although insertion sequences are by no means a universal feature of nuclear 28S genes (they are confined so far to the higher diptera of metazoan animals and to ciliate protozoa), they have been shown to contain sufficient residual interspecific homology to suggest that they are of ancient origin and arose before the divergence of the dipteran

suborder Cyclorrhapha into the calyptrates (such as the house-fly) and acalyptrates (such as *Drosophila*) approximately 65 Myr ago¹¹. Bearing in mind their age it is clear that if complete rDNA arrays are subjected to a concerted 'spring-clean', then the INS⁺ units are either selectively maintained or periodically re-inserted into newly amplified units. For the latter they would require a degree of mobility.

The data of Glover and Rae and their colleagues^{1,3} on the nucleotides surrounding INS elements in *Drosophila* elucidate this question of mobility. In *Drosophila virilis* there is a 14-base pair duplication of a 28S gene sequence that flanks the insert. This same sequence is present only once in INS⁻ units. The generation of a host sequence duplication (usually of 5–9 base pairs) upon insertion of an extraneous sequence is a common feature of mobile elements in *Drosophila* and yeast^{12,13}. In *D. melanogaster* the type I INS is flanked on its right by this same sequence and on the left there is a 9-base pair deletion in the 28S gene. The two inserts clearly slot into identical positions in the two species suggesting a requirement for some sequence homology between host and insert that is analogous to the integration of λ phages into *Escherichia coli*. These data underline the possibility of mobility of INS elements. It is unlikely that there are any fundamental differences between the units that receive INS and those that do not. For example, Long, Rebert and David¹⁴ find no differences between INS⁺ and INS⁻ units in 470 nucleotides around the transcription initiation sites and in the sensitivity to DNase I of the chromatin of genes, spacers and inserts. Is there, then, a limiting supply

Gabriel Dover and Enrico Coen are in the Department of Genetics, University of Cambridge.

of INS elements to re-occupy the 60 per cent of X chromosome 28S genes with type I and the 15 per cent of X and Y 28S genes with type II, or are the numbers selectively constrained? Interestingly, extra copies of type I sequences have been located outside of the NOs in both *D. melanogaster*⁴ and *D. virilis*³ that are situated for the most part in the centromeric heterochromatin in *D. melanogaster*⁵. However, type I INS units in the heterochromatin are each flanked by the same 28S gene segments that flank type I INS in the 28S gene¹, suggesting a transposition of INS with its flanking sequences from the NO to the heterochromatin and not vice versa.

The phenomenon of concerted rDNA evolution in *Drosophila* species has only recently come to light. Electron microscope observations on the extent of heteroduplex matching of interspecific NT spacers of sibling species of the *D. melanogaster* species subgroup initially underscored, at this level of resolution, the conservation, not the dissimilarities, in spacers between the species¹⁵. However, our analysis of the distribution of restriction sites and sequence homologies of NTS, ITS and INS elements of the species nevertheless shows that all these components are divergent between species and yet have evolved in concert within each species to the extent that a complete replacement of one sequence variant by another has occurred in both X and Y NOs¹⁶. Similarly, the NTS sequences of two more distantly related species, *D. melanogaster* and *D. hydei*, have been found not to be the same¹⁷.

Although evidence for the precise mechanism responsible for within-species homogeneity is not available, there are observations that strongly hint at one possible cause. A variety of studies have shown that an heterogeneity in length and an internal repetition within NT spacers are common features of multiple-copy genes (see refs 7 and 18 for reviews). For example, the lengths of NT spacers both within and between sibling species of *D. melanogaster* often differ by multiples of 250 base pairs¹⁶ which is the periodicity of the *AluI* repetition in the NTS of *D. melanogaster* (see figure). It has been suggested^{7,8} that a process of unequal exchange within NT spacers, in which the inequality is of the same periodicity as the internal repetition, would generate the observed differences. Given such a mechanism plus the ability of sequences of DNA to transpose or be exchanged between chromosomes, it is possible to offer some explanation for the complete horizontal replacement, at several different levels, of the rDNA within and between individuals in one group of species.

First, within the NO of one chromosome it is possible to view the replacement of repeats, one variant by another, as an outcome of a continuous process of amplification and elimination of sequences (by unequal exchange) in a stochastic

manner. At the lowest level there would be the replacement of units of the internal repetition within the NTS; for example, in *D. melanogaster* a replacement by a variant repeat (containing an *AluI* site) has taken place in an array of 250-base pair repeats (lacking the *AluI* site) that is probably common to the NT spacers of the sibling subgroup¹⁶. An analogous 250-base pair periodicity has been detected in the NTS of *D. hydei*¹⁷. In addition, unequal exchange within the internal repetition would also give rise to the 250-base pair length heterogeneity described above. At a higher level unequal exchange involving complete rDNA units would lead to the replacement of complete arrays of NT spacers with their variant internal repeats in any one NO.

The second step in the process of sequence encroachment would be the establishment of the same spacer variant in all NOs of a karyotype. N. Arnheim and his colleagues report the complete distribution of spacer variants to several NOs within any one species of primate², analogous to the distribution of *AluI* variants to the X and Y NOs of *D. melanogaster*¹⁶. Such a mobility between chromosomes could be either a direct transposition or dependent on both meiotic (homologous) and mitotic (non-homologous) chromosome exchanges. Finally, and importantly, if the processes of unequal exchange and interchromosome spreading are rapid, then a variant will be driven throughout a population of sexually reproducing individuals: the last stage of concerted evolution. The absence of much rDNA sequence polymorphism, as distinct from length heterogeneity, within species suggests that the mutation rate producing new variants might be low relative to the rates at which they displace and spread both within and between individuals.

Against such a background of periodic replacement of rDNA units, in *Drosophila*, the exclusion of specific INSs from the NOs of the Y chromosomes of *D. melanogaster*⁶ and *D. hydei*⁷ suggests that insertion is a separate and independent stage in the overall evolution of the rDNA clusters.

The role of selection in the evolution of rDNA is much harder to gauge. There is probably a selective constraint on the number of 28S genes that can accept an insertion given the small deletion¹ (Glover

op. cit.) and the absence of transcription of INS⁺ genes in somatic tissues of *D. melanogaster*^{4,19}; although the role of selection is ambiguous in strains of *Tetrahymena pigmentosa* which differ by the total absence or presence of insertions throughout the 28S genes¹¹. There might also be a selective constraint that restricts the length heterogeneity of NTSs to within the defined lengths observed between sibling species of *Drosophila*¹⁶ and within individually characterized X and Y NOs of *D. melanogaster*^{4,16,20,21}. This might ensure the correct phasing of nucleosomes vis-à-vis the transcription initiation sites of the repeats. An interesting regularity in nucleosome phasing with this property has recently been revealed in multiple-copy families of histone and ribosome genes in *D. melanogaster*²² and in 5S RNA arrays of *X. laevis*²³.

One final consideration concerns the generality of the phenomenon of concerted evolution and repeat unit replacement in eukaryote genomes. It is known that repeated sequences are undergoing recurrent rounds of amplification and that on occasion a variant member becomes the fortuitous unit of amplification leading to the establishment of an homogeneous but variant array (segment) (see ref. 24 for references). Such polymorphic segments might represent a state of transition during replacement. Segmental variants and concerted replacement have been observed in tandem arrays of satellite DNAs, histones, immunoglobulins and globins, in addition to rDNA. Recently, evidence for concerted replacement has been found in a set of repeats that are dispersed throughout the genomes of different rodent species²⁵. This might result from a progressive 'gene conversion' of one interspersed repeat by another (rather than unequal exchange) along the lines described by Scherer and Davis in yeast²⁶.

The genomic consequences of sequence replacement, mobility and interchromosome spread, occurring simultaneously, for example, in the rDNA repeats, are clear: the rapid establishment of a group of genomes (species) that are uniquely identified by many shared homogeneous families of sequences. The biological consequences of the accidental fixation of multiple genomic dissimilarities between groups are endless. □

- Rojia, H., Miller, J.R., Woods, L.C. & Glover, D.M. *Nature* **290**, 749 (1981).
- Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O., and Zimmer, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7323 (1980).
- Rae, P.M.M., Kohorn, B.D. & Wade, R.P. *Nucleic Acids Res.* **8**, 3491 (1980).
- Dawid, I.B., Lauth, M. & Wellauer, P.K. *Miami Winter Symp.* **17**, 217 (1980).
- Peacock, W.J., Appels, R., Endow, S. & Glover, D. *Genet. Res. Camb.* **37**, 209 (1981).
- Long, E.O. & Dawid, I.B. *Rev. Biochem.* **49**, 727 (1980).
- Brown, D.D., Wensink, P.C. & Jordan, E. *J. molec. Biol.* **63**, 57 (1972).
- Wellauer, P.K. & Reeder, R.H. *J. molec. Biol.* **94**, 151 (1975).
- Gourse, R.L. & Gerbi, S.A. *Nature* **283**, 693 (1980).
- Wild, M.A. & Sommer, R. *Nature* **283**, 693 (1980).
- Barnett, T. & Rae, P.M.M. *Cell* **16**, 763 (1979).
- Dunsmuir, P., Brorine, W.J., Simon, M.A. & Rubin, G.M. *Cell* **21**, 575 (1980).
- Farabaugh, P.J. & Fink, G.R. *Nature* **286**, 352 (1980).
- Long, E.O., Rebbert, M.L. & Dawid, I.B. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Tartof, K.D. *Cell* **17**, 607 (1979).
- Coen, E. & Dover, G.A. (in preparation).
- Renkawitz-Pohl, R., Glatzer, K.H. & Kunz, W. *Nucleic Acids Res.* **8**, 4593 (1981).
- Fedoroff, N.V. *Cell* **16**, 697 (1979).
- Jolly, D.J. & Thomas, C.A. *Nucleic Acids Res.* **8**, 67 (1980).
- Indik, Z.K. & Tartof, K.D. *Nature* **284**, 477 (1980).
- Endow, S.A. *Cell* **22**, 149 (1980).
- Samal, B., Worcel, A., Louis, C. & Schedl, P. *Cell* **23**, 401 (1981).
- Gottesfeld, J.M. & Bloomer, L.S. *Cell* **21**, 751 (1980).
- Dover, G.A. *Nature* **285**, 618 (1980).
- Brown, S.D.M. & Dover, G.A. *J. molec. Biol.* (submitted).
- Scherer, S. & Davis, R.W. *Science* **209**, 1380 (1980).

Continental margin processes

from David G. Roberts

OVER the last decade research on continental margins has been greatly stimulated by the need to assess the hydrocarbon potential of the largely unexplored margin areas and to develop a thorough geological and geophysical understanding of the transition across the fundamental crustal boundaries defined by passive and active margins. The recent Hedberg Conference* was the third to review continental margins since the original Penrose Conference of 1972 (see *The Geology of Continental Margins* Springer Verlag, 1975).

Results obtained from seismic exploration allied to the International Phase of Ocean Drilling (IPOD) were reviewed by A.W. Bally (Shell Oil, USA). Specific problems of passive margins revealed by the different structural styles of the north Biscay and eastern North American margins include the mechanism and amount of stretching during rifting and its effect on the subsequent history and thermal subsidence. In the case of active shear and convergent margins the structural implications of brittle-ductile boundaries in the lower crust have yet to be fully understood as has the role of crustal extension in an active margin setting. Bally urged full documentation of the palaeontological data base used to erect the global sea-level curves of Vail *et al.* (*Mem. Am. Ass. petrol. Geol.* 26; 44, 1981) and suggested changes in the subsidence rate of margins due to global sea-level variations as an alternative mechanism.

Convergent margins were reviewed by Von Huene (USGS, Menlo Park). IPOD drilling and associated seismic surveys have revealed evidence of uplift and tectonic erosion that have radical implications for mechanisms of thrusting and subduction. Subduction erosion or underplating has been proposed to account for tectonic erosion off Central America while uplift on the Japanese active margin may result from development of a normal, that is, more elevated geothermal gradient during periods of inactive subduction. Recession of the continental edge by as much as 60 km from the trenches may now be accounted for by dewatering and temperature-controlled 'clay' mineral transformations with associated faulting in response to weak shear stress conditions in the accreting sediments.

Other active margins discussed included the Nankai Trough off Japan (Aoki, Japan Petroleum) and the Makran margin where uplift related to convergence is about

1.5–2 km per million years (White, University of Cambridge). Examples from the Caribbean included the margins of Venezuela and Hispaniola where studies of seismic reflection profiles offshore have been integrated with a field study of the exposed parts of the active margin onshore (Biju-Duval, IFP, France). Along the Indonesian Arc (Curry, Scripps Institution) variations in structural style are reflected in the development of strike-slip faults and spreading back arc basins in turn related to the direction of convergence. Episodicity in the geological history of the exposed parts of the Arc may be more apparent than real and is perhaps caused by changes in plate convergence. Uplift and subsidence of part of the Aleutian Arc were also documented by Scholl (USGS, Menlo Park).

The Caledonian–Appalachian orogenic belt was the subject of several interesting interpretative comparisons with the marine results. Leggett (Imperial College, London) demonstrated the particularly striking resemblance between the Southern Uplands of Scotland, the Makran and the Middle American Trench while McKerron (University of Oxford) reviewed the history of the northwestern edge of the Iapetus Ocean. Evidence of very large-scale tectonic transport along convergent margins was documented from California and the Appalachian belt. 'Suspect' terranes in California show 30–40° of rotation and transport of Upper Cretaceous rocks by as much as 3,000 km (Howell, USGS, Menlo Park). COCORP results in Appalachia also demonstrate thrusting of several hundred kilometres along the Blue Ridge thrust and have wider implications in interpreting autochthonous terranes in the Caledonide–Appalachian belt.

The session on passive margin evolution included reports from all over the world. Grachev (Institute Physics of the Earth, USSR) reviewed the Moma rift — the onshore continuation in Siberia of the spreading Gakkel Ridge of the Arctic Ocean. Nunns (Durham, UK) discussed the evolution of the Norwegian Sea and the Jan Mayen microcontinent. The structure of the entire Southwest African margin from the Walvis Ridge to Capetown was presented by Gerard (SOEKOR, South Africa) using seismic reflection and commercial results. This margin, like the Voring Plateau, East Greenland and some other margins, is characterized by a thick sequence of oceanward dipping reflectors close to the continent–ocean boundary. Drilling suggests that these reflections may be basalts. Association of supposedly oceanic magnetic anomalies with

continental crust on this margin questions their use in defining the continent–ocean transition. The Congo–Cabinda area further north is a classical area of rift tectonics where the entire transition from the pre-rift phase through rifting, initiated in the late Jurassic, to the Albian post-rift and the important role of halokines is well documented from seismic and drilling results (Cochran *et al.* Gulf Oil, USA).

Off eastern North America, improved seismic data and the results of recent IPOD legs have shed new light on the evolution of the margins. In the Straits of Florida (Buller and Schwab, UTMSI) drilling on partly exposed tilted blocks has cored gneisses and phyllites. Careful analysis of refraction (Ibrahim and Uchupi, WHOI) and gravity and magnetic data (Hall, Gulf Oil) from the Gulf of Mexico suggest rifting and spreading formed a rhomboidal oceanic in early Jurassic time. Independent evidence for early Jurassic opening of the western North Atlantic has been given by the results of DSDP hole H534 which cored basement of Callovian age east of the Blake Escarpment (Sheridan, Gradstein *et al.*, University of Delaware). In the nearby Carolina Trough, the East Coast magnetic anomaly is interpreted as an edge effect associated with the continent–ocean boundary by Hutchinson *et al.* (USGS–WHOI) who also find no evidence of a basement ridge along the boundary. They propose a model of rapid stretching and subsidence to account for a zone of attenuated continental crust landwards of the boundary. Present growth faulting on this margin is apparently related to tectonically controlled halokinesis (Dillon, USGS, WHOI).

Models of stretching and subsidence provided one of the main points of formal and informal discussion. McKenzie (*Earth planet. Sci. Lett.* 40; 25, 1978) originally proposed a simple model of lithospheric stretching and applications of this model were discussed by several speakers. Sobolev and Artemjev (Institute Physics of the Earth, USSR) discussed rifting and stretching in terms of phase transitions within the mantle. Variations in the width of passive margins may be related to eclogitization. Keen *et al.* (Bedford Institute, Canada) analysed the development of sedimentary basins off North America in terms of the McKenzie model, models of uniform extension with melt segregation and of depth-dependent extension requiring rheological zonation within the lithosphere. They compared calculated with observed gravity, stratigraphy, mantle depth and thermal history data and concluded that the depth-dependent extension model gave the best fit. A rather different view was proposed by Sibuet and Le Pichon whose

David G. Roberts is a Principal Scientific Officer at the Institute of Oceanographic Sciences, UK.

*The Hedberg Conference was held from January 12 to 16, 1981 and was convened to honour Hollis D. Hedberg, Professor of Geology at Princeton and former chief geologist of Gulf Oil. The meeting was sponsored by the American Association of Petroleum Geologists.

calculations of the crustal extension in Biscay were consistent with the uniform stretching model of McKenzie and could be applied to other areas. Chenet *et al.* (IFP, Paris, France) also calculated the extension for the north margin of Biscay and Galicia Bank and calculated that a two-layer model best fitted the data although a mass balance problem remained in the lower crust. Sclater *et al.* (MIT) reviewed subsidence of the Georges Bank in terms of the McKenzie model and discussed possible hydrocarbon potential in terms of the thermal history. A two-layer finite element crustal model used to model the formation of outer highs by Vierbuchen and George (Exon, Houston) included variables such as the temperature at the base of the lithosphere and time-dependent creep. Watts (Lamont-Doherty, New York, USA) reviewed the roles of eustasy, flexural strength, diagenesis, stretching and cooling in margin subsidence, concluding that at least off North America the paucity of well data and the unknown thickness of syn-rift sediments do not easily allow test of McKenzie's model.

Sedimentary processes in many continental margin environments were documented using a variety of techniques. Coumes (SNPA, Pau, France) examined the depositional history of the Cap Ferret Fan, south-west Biscay using Seabeam and high-resolution seismic data linked to cores. A wider view of continental margin sedimentary processes was given by Roberts (IOS, UK) from the results of long-range sonar (GLORIA) surveys. Complete sonar mosaics of entire continental margins are giving a needed overview of the role of downslope and slope sediment movement in determining facies distribution. The role of slope parallel and downslope sediment transport in the development of the lower continental rise off the eastern USA was documented using seismic profiles (Laine and Tucholke, WHOI). A detailed study of the McKenzie River delta was presented by Cole (Gulf, Canada) who recognized five deltaic cycles in the delta's development. Bouma (USGS, Corpus Christi) discussed the influence of halokinesis on upper slope sedimentation in the Gulf of Mexico.

The maturity and characteristics of organic matter on continental margins was appropriately discussed in the final session. Arthur (USGS, Denver) discussed the role of Mesozoic–Cenozoic anoxic events determining the production and presentation of organic matter along the slope and rise. Accumulations are episodic and themselves variable in organic matter and type. A detailed geochemical study of Cretaceous organic matter by Katz (Texaco, USA) showed differences between the North and South Atlantic Oceans. The North Atlantic organic matter is predominantly of terrestrial origin and is likely to yield gas compared with the oil potential of the marine organic matter in the South Atlantic Ocean. □

Solar physics at Taos

from R. Rosner

THE second major presentation of results from the Solar Maximum Mission (SMM) satellite, which recently met an untimely failure in its pointing system after approximately six months of active operation, provided the focal point at the winter meeting of the American Astronomical Society's Solar Physics Division*. This gave the solar community an excellent overview of both the current status and the likely scientific impact of the SMM venture. Several other major results not associated with the SMM programme were also discussed.

One of the first new results reported concerns periodic fluctuations in hard (≥ 20 keV) X-ray emission from solar flares. Such fluctuations have been observed by both the SMM Hard X-Ray Burst Spectrometer (HXRBS), reported by Kiplinger (Goddard Space Flight Center), and an X-ray spectrometer on board the ISEE-3 spacecraft (Kane, UC Berkeley, who also reported corresponding quasi-periodic variations in microwave radio bursts). The typical pulse period is ~ 8 s, on a coherence time scale of the order of several minutes. Several explanations were offered for these observations; one plausible model proposed by Morrison and Ionson (Goddard) considers the flaring 'loop' structure as a lumped resonant circuit whose solutions on short time scales successfully predict oscillatory behaviour with periods of the order of seconds.

Detailed flare plasma diagnostics have been considerably advanced by SMM observations. High-quality observational evidence for both systematic flows and turbulent velocities were reported by the Ultraviolet Spectrometer and Polarimeter (UVSP) and X-Ray Polychromator (XRP) experiment teams; for example, turbulent velocities (as measured by broadening of lines due to high ionization states of calcium and iron) were shown to attain values up to ~ 100 km s⁻¹ during flares. Data from these two instruments and from the HXRBS have been assembled by Frost, Orwig, Poland and Shine (Goddard) who reported evidence that considerable hard X-ray flare emission originates from the vicinity of flare loop footpoints. These data, together with high-resolution Very Large Array (VLA) observations (Kundu, University of Maryland, and collaborators), strongly support the view that a non-thermal component is present in the electron velocity distribution and is responsible for

the footpoint hard X-ray emission. Yet further evidence for non-thermal processes comes from the Gamma-Ray Spectrometer (GRS): Forrest (University of New Hampshire), Ramaty (Goddard) and collaborators showed spectra which clearly indicated line emission at, for example, 2.2 MeV (attributable to neutron capture by protons) and 0.51 MeV (corresponding to the well known e^-e^+ annihilation line), and demonstrated convincingly that both ions and electrons must be accelerated impulsively (and episodically) during a flare on similar time scales; the detailed temporal variation of the 2.2-MeV line flux was shown to follow the hard X-ray flux variation and to correspond to the behaviour expected from neutron slow-down and eventual capture. In a review of the high-energy experimental flare results from SMM, P. Hoyng (Utrecht) argued that the preponderance of evidence now points towards the 'non-thermal' flare models (although, as pointed out by A. G. Emslie (Stanford), spatial characteristics of hard X-ray emission which may be attributed to a non-thermal model can also be predicted by some 'thermal' models).

The considerable flare emphasis of SMM may lead one to forget that one instrument aboard SMM was especially designed to observe phenomena on longer time scales — ACRIM, a cluster of three active cavity radiometers designed to monitor the possible variability of the solar bolometric luminosity. This important, but controversial subject was reviewed by S. Sofia (Goddard), and Hudson (UC San Diego), Woodard (UC San Diego) and Willson (JPL) reported luminosity fluctuations as large as ~ 0.2 per cent (probably associated with the passage of sunspot groups across the solar disk). ACRIM should help to answer some of the outstanding questions in this area; for example, it can now be shown observationally that the well known sunspot energy 'deficit' is not spatially redistributed in the vicinity of a spot on short time scales, suggesting that transient energy storage does indeed occur in the solar convection zone.

There is, of course, an enormous amount of solar physics conducted outside the SMM realm. J. Kohl (Harvard-Smithsonian CfA) and collaborators are using resonantly scattered Ly α photons as a diagnostic tool for coronal plasma above 1.5 solar radii. They have had two successful rocket flights, and G. Withbroe (Harvard-Smithsonian CfA) reported one of the first major results — that proton and

*The meeting of the Solar Physics Division of the American Astronomical Society was held 7–10 January 1981 at Taos, New Mexico.

R. Rosner is at the Harvard-Smithsonian Center for Astrophysics.

electron temperatures rapidly decouple above ~ 2 solar radii. The relative spatial constancy of the electron temperature, in contrast to the rapidly declining proton temperature with increasing radius, implies that the coronal heating process primarily couples to electrons, and is consistent with no significant proton heating (other than via e^- - p coupling) above ~ 2 solar radii. IMP satellite data (Borini, Stanford, and Gosling, Los Alamos) included the first definitive identification of (at least one of) the coronal sources of the low-speed solar wind, that is, the 'belt' of coronal streamers overlying photospheric neutral lines, and anomalously low He III abundances in the high plasma density wind regions associated with sector boundaries and (when mapped back to the Sun) coronal streamers. It is still debatable

whether yet other coronal source regions of the low-speed solar wind exist; these IMP results (particularly the He abundance anomaly), however, provide an important constraint for solar wind theorists.

Although experimenters did rather steal the show, theorists too were present at the meeting. Amongst the interesting theoretical results reported were the hydrodynamic calculations of the chromosphere, transition region and corona (McClymont and Canfield, UC San Diego; Boris and Mariska, NRL; and Pallavicini, Arcetri Observatory) which promise to resolve the controversy on the structure and stability of the 'confined' solar atmosphere, and convection zone calculations (Hurlburt and Toomre, JILA) which may help understanding of convection in the fully compressible limit. $\square$

Neutron optical activity

from Leo Stodolsky

THE PHENOMENON of optical activity, in which a substance rotates the plane of polarization of light passing through it, is a familiar one. It is a common property of organic chemicals and is often used, for example, to measure their concentrations. Recently, a group working at the ILL (Institut Laue-Langevin) in Grenoble has reported¹ the first observation of the equivalent phenomenon for matter waves. Furthermore, the size of effect appears, in at least one case, to greatly exceed simple theoretical expectations. In an experiment using cold neutrons from the ILL research reactor (see the figure), Forte and his colleagues have seen the rotation of the spin of the neutron around its line of flight as it propagates in solid tin. A neutron with its spin direction perpendicular to its motion is a linear combination of two helicity (or circular polarization) states, just like a linearly polarized photon; therefore such a neutron beam is exactly analogous to plane-polarized light.

The neutron spin rotation observed differs from the optical activity of chemicals in one fundamental respect, however. In both effects a rotation (of the spin or plane of polarization) is coupled with a direction (line of flight) and demonstrates a 'handedness' or parity sense associated with the material. In the case of ordinary optical activity this handedness is traceable to the physical structure of the molecules of the material, which are in a configuration of definite handedness (stereoisomerism). A sample of tin, however, contains no such handed structures. (The possibility of a neutron spin rotation created by handed materials,

through nuclear forces², or through magnetic interactions³, has been considered, but the predictions are quite small.) Thus the parity sense manifested by the neutron beam must be traced ultimately to a handedness in the interaction forces between the neutron and the constituents of the material — that is, to the phenomenon of parity violation in the fundamental interactions. Thus the small (10^{-6} rad) but definite rotation measured is a striking manifestation of a parity sense imparted to bulk matter by parity-violating forces.

The possibility of such a 'neutron activity' was, in fact, first suggested⁴ by Michel in 1964 as a way of seeing the parity-violating, weak interaction forces between the neutron and the atomic nucleus. The interest of the effect in connection with the question of 'neutral weak currents' and the possibility of observing the neutron-electron weak interaction was recognized seven years ago⁵ and brought to the attention of Norman Ramsey (Harvard University), who has been involved in a long series of experiments searching for a

neutron electric dipole moment, where such very small neutron spin rotations must be measured.

The present experiment, an Ispra/Harvard/Rutherford/Sussex collaboration, may be viewed as a neutron version of the standard crossed-polarizer configuration used in optics. Polarized neutrons pass from a polarizer through the tin samples, which are about 5 cm long, to a symmetric analyser, which yields neutron counts depending on the degree of spin rotation. Care must be taken to understand and eliminate magnetic field effects which could give spurious rotations by acting on the neutron magnetic moment, and various control and null tests were performed.

In classical fashion, the effect first showed up not in the material being studied (^{124}Sn) but in a control sample (natural tin) from which a large effect was finally located in ^{117}Sn .

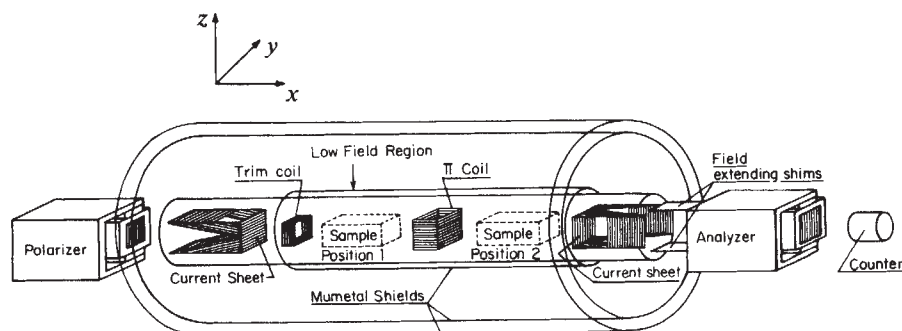
Results for the three materials are:

^{124}Sn , $(0.48 \pm 1.49) \times 10^{-6}$ rad cm^{-1} ;
 ^{117}Sn , $(36.7 \pm 2.7) \times 10^{-6}$ rad cm^{-1} ;
 $\text{Sn}(\text{natural})$, $(4.95 \pm 0.93) \times 10^{-6}$ rad cm^{-1} ;
 Since natural tin contains only 7.6 per cent ^{117}Sn , there would appear to be a residual effect of $(2.16 \pm 0.95) \times 10^{-6}$ rad cm^{-1} in natural tin not explained by its ^{117}Sn content. It should be noted that large parity-violating effects have shown up in ^{117}Sn before, in neutron radiative capture⁶.

In the context of the present picture of weak interactions, estimation of the size of the spin rotation effect would seem to be straightforward⁵. The neutron-atom weak interaction contains a term $\sigma \cdot p$, proportional to the helicity of the neutron, which creates a different index of refraction for each of the two helicity states of the neutron in matter. This in turn induces a rotation of the spin component perpendicular to the motion. The rotation in radians, Φ , for a neutron travelling a distance Z in a simple Born approximation calculation is:

$$\Phi = \sqrt{2} G W \rho Z$$

where G is the Fermi weak interaction constant, ρ the number density of atoms and W a 'weak charge' for the atom. W is expected to be on the order of the number of particles in the atom, about 70 for example in the currently accepted electroweak theory of weak interactions. Among the



Schematic view of experimental apparatus. The neutron beam polarization is along z and momentum along x .

Leo Stodolsky is at the Max-Planck Institute for Physics and Astrophysics, Munich.

interesting features of the formula are that it is proportional simply to G and that, in contrast to ordinary optical activity, which must vanish at low frequency, it is energy independent. The observation of the effect with such low-energy (millivolt) neutrons appears to verify this last feature.

The size of the observed effect is surprising, however. The above theoretical formula for Φ , as well as more sophisticated calculations (for discussion and references, see ref. 7), lead to an estimate of about 10^{-8} rad cm^{-1} . This is 3,000 times smaller than that observed for ^{117}Sn .

The only explanation so far put forward to explain such a large discrepancy involves a very low-energy neutron (P-wave) resonance. In fact, the original motivation⁸ for studying ^{124}Sn — where, ironically, no effect at the 10^{-6} rad cm^{-1} level was found — was the suggestion of a resonant enhancement for the effect. In ^{117}Sn a weak resonance, which could be P wave, has been reported (1.2 eV). Even with the existence of such a resonance, however, it seems difficult to get an effect of the observed magnitude. This has prompted at least the present author to wonder whether there may not be some surprises lurking in the still not very well understood sector of the fundamental interactions concerning the non-leptonic weak interactions, particularly the weak interactions between nucleons⁸. The imperative question here is whether the large rotation is indeed some kind of special nuclear physics effect particular to ^{117}Sn and perhaps a few other rare nuclides, or whether there is something unexpected generally residing in the nucleon–nucleon force. In this respect the residual rotation for natural tin, beyond that expected from its ^{117}Sn content, is intriguing. If true, it indicates a large rotation on some other isotope also, and makes it much less likely that what is happening with ^{117}Sn is a freak occurrence. These questions can be clarified by measurements on a variety of nuclei, particularly the lighter ones, where the resonances are widely spaced and well known. Should the effect consistently turn out to be too large, it could point to the need to revise our understanding of the weak interactions on a basic level.

Finally, another interesting fact has emerged from this experiment. Motivated by the large rotation on ^{117}Sn , Forte and his co-workers looked for a dependence of the cross-section on the helicity of the neutron, that is, a difference in the rate of

neutrons transmitted through the material according to whether the spin is parallel or antiparallel to the direction of motion. This is also manifestly parity violating, and quantum mechanically is governed by (the imaginary part of) the same amplitude as the spin rotation. Here again a large effect is found. Although the authors say the result is too preliminary to be quoted as an actual cross-section, it is stated as being at about 9×10^{-6} of the total cross-section. The magnitude of this may be appreciated by saying that it corresponds to a cross-section at the tens or hundreds of μb level, a number more typical of say, electromagnetic than weak interactions, and hence appears 'big'.

This observation is something of an embarrassment to the theoretical workers in the subject. Although we had considered

the polarization dependence of the cross-section, the possibility that it could be finite, let alone 'big', at threshold was never considered since it would appear to involve inevitably a P-wave suppression factor. A finite parity-violating cross-section at zero energy is, in fact, possible, however, when an exothermic channel such as radiative capture or perhaps fission is present. This leads to an amusing and complex situation where, for example, the dominant process for the P-wave scattering is parity violating. Furthermore, because of the spin structure, some interesting tests of time reversal invariance should be possible in this novel system.

Although it is not at all clear that this situation is fully understood, it does seem clear that the door has been opened here to a new set of very interesting phenomena. □

Two DNA-binding proteins

from David Davies

THE last few years have seen dramatic developments in our view of DNA as a dynamic molecule, both in structure and function. Most of these properties can be attributed to the proteins that interact with DNA to produce replication, transcription, translation, transposition and processing. There are polymerases, synthetases, untwisting proteins, twisting proteins, histones, repressors, activators. . . , but until recently very little was known about their three-dimensional structure or how, precisely, they interact with nucleic acid. This has now begun to change with the X-ray diffraction studies reported in this issue of *Nature* on two DNA-binding proteins, the *cro* repressor of bacteriophage λ (p.754) and the *E. coli* catabolite gene activating protein (CAP) (p.744). Although neither crystal structure actually involves the protein–DNA complex, the protein structures in both cases immediately suggest how the complex might form. The remarkable conclusion is made that while *cro* binds to right-handed B-DNA, CAP binds to left-handed B-DNA.

Cro is a small protein (molecular weight 7,351) that, together with the λ repressor protein, acts in the control of the lytic pathway of the bacteriophage³. The structure of the monomer at 2.8 Å resolution, as determined by Anderson *et al.*¹, is very simple, consisting of three extended segments forming an antiparallel β sheet, together with three helices. An

extended C-terminal tail interacts strongly with the corresponding tail of another monomer to form a dimer. This dimer associates in the crystal with another such dimer to form a tetramer, although in solution it is not known whether *cro* is a dimer or tetramer. However it is known that it binds to three sites on λ DNA, each of which exhibits some two-fold (palindromic) symmetry suggesting that the interacting form of the protein will be at least a dimer with two-fold symmetry.

Insight into the structure of the *cro*–DNA complex has come from a combination of genetics, chemistry and model building. *Cro* binds to a 17-nucleotide length of the DNA and prevents alkylation of the phosphates and methylation of the guanines at a site in the major groove. *Cro* appears to bind principally in the major groove of the DNA since methylation of adenine in the minor groove is not prevented by binding.

What parts of the *cro* structure could be involved in binding to DNA? One very suggestive feature is the helix from G1n 27 to Ala 36, which protrudes from the surface of the protein. The corresponding helix in the other subunit of the dimer is 34 Å away and each helix is inclined at 32° to the line joining their midpoints. These are the dimensions of the grooves on B-DNA and model building indicates that the helices can be positioned to fit convincingly into the major groove of right-handed B-DNA. Between the helices, some residues of the C-terminal segment form a pair of antiparallel β -sheet strands parallel to the minor groove which may interact

1. Forte, M. *et al.* *Phys. Rev. Lett.* **45**, 2088 (1980).
2. Kabir, P.K., Karl, G. & Obyrk, E. *Phys. Rev.* **D10**, 1471 (1970).
3. Harris, R.A. & Stodolsky, L. *J. chem. Phys.* **70**, 2789 (1979).
4. Michel, F. C. *Phys. Rev.* **133**, 3329 (1964).
5. Stodolsky, L. *Phys. Lett* **50B**, 352 (1974).
6. Benkoulal, H. *et al.* *Phys. Lett.* **71B**, 287 (1977).
7. Karl, G. in *Fundamental Physics with Reactor Neutrons and Neutrinos* (ed. von Egidy, T.) (Institute of Physics, London, 1977).
8. Forte, M. in *Fundamental Physics with Reactor Neutrons and Neutrinos* (ed. von Egidy, T.) (Institute of Physics, London, 1977).
9. Stodolsky, L. *Phys. Lett.* **96B**, 127 (1980).

David Davies is in the National Institutes of Health, Bethesda, Maryland.

with it.

Another regulatory protein, CAP, forms a complex with cyclic AMP and binds to specific DNA sites thereby altering the rate of transcription. The structure of the complex CAP-cAMP has been determined by McKay and Steitz at 2.9 Å resolution. The molecule is a dimer. Each subunit contains two domains joined by a 4–5 amino acid hinge region. There is a cleft between the two domains and in one subunit this cleft is open while in the other it appears closed, implying flexibility of the hinge analogous to that observed in immunoglobulins. The larger N-terminal domain interacts with the corresponding domain of another subunit to form a dimer in which there is local two-fold symmetry between the domains. There appears to be no interaction between the two smaller C-terminal domains.

As in the *cro* protein, the obvious candidates for binding to the DNA are two protruding helices; one from each of the C-terminal domains. However, unlike *cro*, these helices are oriented in such a way that they cannot fit into the grooves of right-handed B-DNA but could bind to the major groove of *left-handed* B-DNA. Such a model of the complex can account for the genetic and chemical data in a manner that could not be done with any model involving right-handed DNA.

Supporters of left-handed DNA recently received a boost with the discovery of the 'Z' structure for the alternating copolymer (dG-dC)⁴. However the [CAP-cAMP-DNA] complex proposed by McKay and Steitz involves left-handed B-DNA and not Z-DNA which has quite-different helical dimensions. In 1979 Crick *et al.*⁵ examined the evidence on the handedness of DNA and concluded that there was little doubt that it was right-handed. Strong support for this conclusion has come from the recent dodecanucleotide crystal structure of Wing *et al.*⁶ who showed that the molecule is a right-handed B double helix. A modelling study by Gupta *et al.*⁷ has demonstrated that a left-handed model can be constructed for B-DNA, but experimental evidence

strongly favours the right-handed model.

McKay and Steitz therefore hypothesize that the handedness of the DNA at the binding site must change when it complexes with CAP. In view of the apparent flexibility of the protein hinge and the loose attachment of the DNA-binding domain to the major domain of CAP, this change of hand is a surprising accomplishment. McKay and Steitz have examined the possibility of rearranging the relative locations of the helices, for example by rotation about the hinge, but conclude that no simple rearrangement will provide a configuration complementary to right-handed B-DNA. Proof of this model will have to come from unwinding experiments on closed circular DNA, or from the crystallization of the CAP protein complexed to a small DNA fragment of appropriate sequence.

Given the complexity of protein structures, it is interesting that both the *cro* and CAP proteins have structures that immediately suggest that their principal means of interaction with the DNA is through α -helices sticking in the major grooves. The only crystallographic analysis of a nucleic acid-protein complex has been

that of protamine-tRNA⁸ which showed an α -helical piece of protamine lying in the minor groove of the tRNA. However, there have been other modelling proposals involving β -structure^{9–11} and the *cro* protein may also utilize the binding of an anti-parallel pair of strands in the minor groove of the DNA. Analysis of further DNA-binding proteins may reveal a more complex pattern of interactions although it should be remembered that DNA is a groovy molecule that might favour the binding of the ordered segments.

Although a model of DNA sequence recognition is not yet revealed by these studies, it is interesting that both proteins are at least dimers and can thus impose some 2-fold symmetry on the DNA binding site. Further information can be expected to come from a refinement and higher-resolution analysis of these and other structures. Also, although the three-dimensional structures of *cro* and CAP are quite acceptable, being based on experimental analysis with a proven technique, the predictions about the complexes are less secure and illustrate the need for crystal structure determinations directly of complexes of protein and DNA.

Laminin and epithelial cell attachment

from Brigid Hogan

THE past few years have seen an increasing interest in epithelial cells and the factors responsible for maintaining their organization into sheets and glands. The subject is important because most human tumours arise within epithelia, and metastasis involves cells breaking away from the territorial boundaries of these tissues and invading the surrounding stroma. Also important is the question of what distinguishes this pathological invasion by malignant carcinoma cells from normal migrations seen during the embryonic development of organs like the breast, when cords of cells burrow from the surface into the underlying mesenchyme to form complex branching ducts and tubes.

Such questions have focused attention on the physical boundary between epithelium and the stroma. In nearly all cases this is an extracellular basement membrane, or basal lamina, made up of proteoglycans, Type IV procollagen, fibronectin, laminin and other glycoproteins. Apart from restraining the wayward migration of epithelial cells by promoting their attachment and

alignment, roles suggested for the basal lamina include a means of laying down specific landmarks for cell guidance, acting as a filter for molecules of a particular size or charge, and the induction of epithelial cell differentiation. Given all this speculation, it is not surprising that the recent identification of the glycoprotein laminin¹ as a major constituent of basement membranes, both in the adult² and the embryo^{3–5}, has generated a whole spate of experiments on the effect of the purified protein on epithelial cells in culture.

Conveniently for the biochemist, laminin is made in large amounts by a number of mouse tumours, including the EHS 'sarcoma' (see ref.1) and several derived from an embryonic tissue known as parietal endoderm⁶. Laminin isolated from the matrix of these tumours consists of a very high molecular weight (approximately 900K), disulphide-bonded complex of three polypeptide chains, a larger one of about 450K and two smaller ones of very similar molecular weight (about 230K). Antibodies to laminin also immunoprecipitate a fourth polypeptide of molecular weight 150K, which is made in much larger amounts by normal rather than tumorigenic endoderm cells⁵, but the role of this laminin-related protein is not yet clear.

Brigid Hogan is in the Mammalian Development Laboratory, Imperial Cancer Research Fund Laboratories, Mill Hill, London.

1. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Matthews, B. W. *Nature* **290**, 754 (1981.)
2. McKay, D. B. and Steitz, T. A. *Nature* **290**, 744 (1981).
3. Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M. and Sauer, R. T. *Cell* **19**, 1 (1980).
4. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. *Nature* **282**, 680 (1979).
5. Crick, F. H. C., Wang, J. C. and Bauer, W. R. *J. molec. Biol.* **129**, 449 (1979).
6. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. *Nature* **287**, 755 (1980).
7. Gupta, G., Bansal, M. and Sasisekharan, V. *Int. J. Biol. Macromols* **2**, 368 (1980).
8. Warrant, r. W. and Kim, S.-H. *Nature* **271**, 130 (1978).
9. Carter, C. and Kraut, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 283 (1974).
10. Church, G. M., Sussman, J. L. and Kim, S.-H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1458-1462 (1977).
11. Blake, C. C. F. and Oatley, S. J. *Nature* **268**, 115 (1977).

There is no doubt, however, that laminin is completely distinct from fibronectin, that other widespread component of basement membranes. In contrast to fibronectin, laminin is only made by epithelial cells and a few others including myotubes and neuroblastoma⁷ and not by fibroblasts. Although it is secreted into the medium of cultured cells, it does not by itself form an extensive fibrous network on both the upper and lower surfaces of monolayer cells, but accumulates as a matrix between the cells and the dish. Laminin does, however, seem to share some properties with fibronectin. Both bind to proteoglycans⁸ and to collagen, but whereas fibronectin binds to all collagens, laminin only seems to interact with Type IV procollagen, the species of collagen which is found in basement membranes. Very little is known about how these properties of laminin influence its deposition *in vivo*. In the skin, the basal lamina is composed of two layers, the *lamina lucida* next to the epidermis and the

lamina densa next to the dermis. Immunofluorescence studies have localized laminin to the *l. lucida*² and Type IV procollagen to the *l. densa*⁹ suggesting that here, at least, there is no intimate intermingling of the two components. Alternatively, laminin may be present throughout the basal lamina but its antigenic sites masked in the *l. densa*.

Although its almost ubiquitous distribution in basement membranes suggests that laminin plays an important role in epithelial tissue organization, is there any real evidence for this? Studies in which epithelial cells were given a choice of collagen matrix on which to attach showed that they bind slowly but preferentially to Type IV procollagen¹⁰. It now appears¹¹ that this binding specifically requires laminin. Normally the cells make it for themselves slowly, but attachment can be speeded up by adding purified laminin, an effect which is blocked by anti-laminin antibody. Laminin also appears to be the active component of various conditioned media which promote epithelial attachment and growth¹². Fibronectin cannot substitute for laminin in these assays, and, conversely, laminin does not mediate the binding of fibroblasts to collagen. These results suggest, but by no means prove, that at least one *in vivo* role of laminin is to help epithelial cells keep their feet firmly on the ground.

It is an important feature of epithelial sheets and tubes that the cells are asymmetrical, with distinct basal and apical surfaces. During embryonic development

there are numerous descriptions of epithelial tissues emerging from clusters of precursor cells which are not themselves polarized in this way, and it is of interest that in two cases the onset of laminin synthesis seems to correlate with the time of this polarization. One example is the morula stage mouse embryo³ which at the time of compaction undergoes transition into a group of cells with definite inner and outer surfaces. The second example is the aggregation of mouse embryo kidney mesenchyme into tubules¹³. In both cases laminin synthesis may only be an early manifestation of epithelial differentiation, but it is also possible that its assembly into a basement membrane actually directs cell polarization in some way.

Finally, what happens to basement membrane laminin when epithelial cells metastasize? In a recent paper, Hayman and colleagues¹⁴ showed that after virus transformation, rat kidney cells lose the ability to lay down laminin and fibronectin in an extracellular matrix, even though they continue to synthesize the two glycoproteins at the same or increased rates. The authors speculate that a matrix cannot be organized by the transformed cells because of a change in their surface glycosaminoglycans, to which, as we have seen, both laminin and fibronectin bind. However, it could also be argued that the primary lesion is in the polarized cytoskeletal framework of the cells which allows them to spread their feet out over a surface, and through transmembrane connections keeps the matrix firmly anchored down. □

1. Timpl *et al.* *J. biol. Chem.* **254**, 9933 (1979).
2. Foidart *et al.* *Lab. Invest.* **42**, 336 (1980).
3. Leivo *et al.* *Dev. Biol.* **79**, 100 (1980).
4. Howe & Salter *Dev. Biol.* **77**, 480 (1980).
5. Hogan *et al.* *Dev. Biol.* **80**, 289 (1980).
6. Chung *et al.* *Cell* **16**, 277 (1979); Sakashita & Rouslahti *Archs Biochem. Biophys.* **205**, 283 (1980).
7. Alitalo *et al.* *Nature* **287**, 465 (1980).
8. Sakashita *et al.* *FEBS Lett.* **116**, 243 (1980).
9. Yacita *et al.* *J. invest. Dermatol.* **70**, 191 (1978).
10. Murray *et al.* *J. Cell Biol.* **80**, 197 (1979).
11. Terranova *et al.* *Cell* **22**, 719 (1980).
12. Vlodavsky & Gossardowicz *Nature* **289**, 304 (1981).
13. Ekblom *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 485 (1980).
14. Hayman *et al.* *J. Cell Biol.* **88**, 352 (1981).



100 years ago

MR. DARWIN ON VIVISECTION

"Dear Sir, — In answer to your courteous letter of April 7 I have no objection to express my opinion with respect to the right of experimenting on loving animals. I have all my life been a strong advocate for humanity to animals, and have done what I could in my writings to enforce this duty. Several years ago, when the agitation against physiologists commenced in England, it was asserted that inhumanity was here practised and useless suffering caused to animals. . . it is right to add that the investigation of the matter by a Royal commission proved that the accusations made against our English physiologists were false. From all that I have heard however I fear that in some parts of Europe little regard is paid to the sufferings of animals, and if this be the case I should be glad to hear of legislation against inhumanity in any such country. On the other hand I know that physiology cannot

possibly progress except by means of experiments on living animals, and I feel the deepest conviction that he who retards the progress of physiology commits a crime against mankind. Any one who remembers, as I can, the state of this science half a century ago must admit that it has made immense progress, and it is now progressing at an ever-increasing rate.

"What improvements in medical practice may be directly attributed to physiological research is a question which can be properly discussed only by those physiologists and medical practitioners who have studied the history of their subjects; but, as far as I can learn, the benefits are already great. However this may be, no one, unless he is grossly ignorant of what science has done for mankind, can entertain any doubt of the incalculable benefits which will hereafter be derived from physiology, not only by man, but by the lower animals. Look, for instance, at Pasteur's results in modifying the germs of the most malignant diseases, from which, as it so happens, animals will in the first place receive more relief than man. Let it be remembered how many lives and what a fearful amount of suffering have been saved by the knowledge gained of parasitic worms through the experiments of Virchow and others on living animals. In the future every one will be astonished at the ingratitude shown, at least in England, to these benefactors of mankind. As for myself, permit me to assure you that I

honour, and shall always honour, every one who advances the noble science of physiology.

"Dear sir, yours faithfully,
"CHARLES DARWIN

Mr. M. G. Mulhall sends us the following curious note, which we give without comment: — "Although Shakespeare is supposed to have taken the idea of Hamlet from the Danish historian Saxo-Græmmaticus, there are such points of resemblance with the Arabic chronicle of Nigiaristan, respecting Montasser, tenth Caliph of Bagdad, that I venture to call your attention to the same. The points of analogy are as follows: 1. That Montasser is murdered by putting poison in his ear. 2. The ghost scene, in which his father appears to him. 3. The displaying of tapestry before the Caliph and his court, in which the tapestry represents a tragedy identical with the late Caliph's murder."

Mr. H. L. Janssen van Raay writes to us from Batavia, March 21, that in the enumeration of the different geographical societies of the world in *Nature*, vol. xxiii. p. 299, the Geographical Society at Samarang (Java), founded in 1879, was omitted.

From *Nature* **23**, 21 & 28 April, 1880.

ARTICLES

The Earth's radius and the G variation

V. M. Canuto

NASA Goddard Space Flight Center, Institute for Space Studies, New York, New York 10025 and
Department of Physics, The City College of the City University of New York, New York 10021, USA

It has long been assumed that if the gravitational constant G was larger in the past, the Earth's radius had to be smaller. The assertion holds provided the input from microphysics (in particular the equation of state) is independent of G . While this is true for some theories of gravity with variable G it is not so in the scale covariant theory, where the pressure can be affected by a variable G in a way that, for a constant mass of the Earth, a larger G in the past implies a larger Earth's radius. Comparison with recent palaeomagnetic data is presented.

MODERN observations¹ of increasing accuracy seem to confirm that the phenomenon of gravity is correctly understood in terms of Einstein's geometrical interpretation and satisfactorily described by the Einstein equations. There, therefore, seems to be no cogent reason to modify Einstein's theory as a description of gravity, as there is no reason to modify atomic theory. Within their limits, gravitational theory and atomic theory are complete and thus far satisfactory theories.

In cosmology, however, where gravitational phenomena involving large fractions of the age of the Universe are analysed not with gravitational but with atomic clocks, the two dynamics must be coupled. Let Δt_E (E for Einstein) be the time interval marked by a clock governed by gravitational dynamics (two orbiting planets, for example) and Δt the time marked by an atomic clock. Because the gravitational constant G and the total mass M may be thought of as the 'springs' of the gravitational clock, much in the same way as e , $\hbar$ and m may be considered the 'springs' of the atomic clock, then $\Delta t_E = f(G, M)$ and $\Delta t = g(e, \hbar, m)$.

To study the large times involved in cosmology a relationship between Δt_E and Δt as a function of time or equivalently of the age of the Universe is required. The lack of a unified theory of electromagnetic and gravitational forces, which would give us the answer, has been avoided so far by assuming that the ratio $\beta(t) = \Delta t_E / \Delta t$ at any time in the past was the same as today. This is known as the strong equivalence principle, SEP. While the adoption of the SEP has not yet led to flagrant disagreement with observations, it remains an unproved assumption, because its confirmation would require two measurements of $\beta(t)$ at two different times, the age of the Universe serving as a time scale. A more popular way of expressing a possible non-constancy of $\beta(t)$ is by saying that G may vary with respect to atomic time². In fact, a non-constant $\beta(t)$ is equivalent to one of the springs of the gravitational clock, say G , varying with respect to the atomic time.

Given the fundamental role of the SEP for the entire cosmological edifice, it is natural to expect that the nature of the function $\beta(t)$ be determined not by an *a priori* assumption, but by either a unified theory, which we do not have, or by observations of which we now have enough, covering a large fraction of the age of the Universe.

Cosmological data to put limits on β/β were used in ref. 3 while the past luminosity of the Sun was analysed in refs 4 and 5. This article concentrates on the dependence of the radius of the Earth on $G = G(t)$.

The newtonian framework is shown to be unsuitable because it leads to inconsistencies (see refs 3, 4, 6, 7). Then it is shown that the Einstein framework leads, as expected, nowhere. Finally, the scale covariant theory is used through which the effects of a non-constant $\beta(t)$ can be treated and quantified consistently.

The Earth's radius

While the first analysis⁸ of the consequences of $G = G(t)$ dealt with the Sun's luminosity (the result $L_\odot \sim G^7$ has been shown⁴ to be $L_\odot \sim \text{constant}$; see also ref. 5), more attention seems to have been given to the Earth's radius^{9,10}. The general consensus seems to be that if G was larger in the past, the Earth's radius had to be smaller (see, for example, refs 9–12).

To arrive at this qualitative conclusion, one uses (a) the hydrostatic equilibrium equation, (b) an equation of state and (c) the assumption of a constant total rest mass, ($\rho_0 = mn$, $n = N/V$, $M \equiv mN$)

$$\frac{1}{\rho_0} \frac{dp}{dr} = -G \frac{m(r)}{r^2} \quad (1a)$$

$$p \propto \rho_0^\gamma \quad (1b)$$

$$M = \text{constant} \quad (1c)$$

thus deriving

$$R \propto M^{(\gamma-2)/(3\gamma-4)} G^{-1/(3\gamma-4)} \sim G^{-1/(3\gamma-4)} \quad (2)$$

which is the basic relation for all the qualitative conclusions that a larger G implies a smaller R , and therefore an expanding Earth.

While equations (1a), (1b) and (1c) form a set of consistent relations when G is constant, is the same true when G becomes a variable? The different exponentials of G and M in equation (2) are in sharp contrast with equation (1a) where G enters only in the combination GM . The asymmetry is introduced using an equation of state in the form (1b), which contains M but not G , the usual justification being that atomic relations are immune from whatever happens to G , which is valid only if gravity and atomic physics are kept separated. However, a variation of G represents a change of one dynamics with respect to the other, a process that can only be meaningful if we relate, not separate, the two dynamics. While the assumption that the two dynamics are disconnected may be logically consistent with the SEP, the purpose here is to relax the SEP allowing the two dynamics to depend on one another through the function $\beta(t)$ or equivalently $G(t)$. Furthermore, while we may reasonably expect an isolated hydrogen atom to be independent of β , this need no longer be true for a many-body quantity like an equation of state, which involves ingredients like the integration over phase space, the Liouville equation, as well as thermodynamic relations.

To show how equation (1b) may lead to incompatibilities with a $G = G(t)$, let us consider the equation of state $pV = NkT$, and use the adiabatic relation $TV^\Gamma = \text{constant}$, where $N = \text{constant}$, that is $M = \text{constant}$. We obtain $p \propto \rho_0^\gamma$ ($\gamma = 1 + \Gamma$), that is equation (1b) (equations (1b) and (1c) are, therefore, related). Furthermore, the basic ingredient in this derivation is the

adiabatic relation $TV^\Gamma = \text{constant}$, which comes from integrating the first law of thermodynamics, which in turn is a statement of conservation of energy.

However, if $G = G(t)$, the potential energy $F = -G(t)m_1m_2/r^2 = -\nabla V(r, t)$ becomes time dependent and we do not have energy conservation¹³. An inconsistency has surfaced.

The argument may be formalized by moving from the newtonian framework to the einsteinian one, not because general relativistic effects are relevant to the Earth, but because GR is the correct description of gravity. While the previous argument indicates that equations (1b) and (1c) are related, adopting the einsteinian framework will show that all the relations (1) are connected. In fact, the derivation of equation (1a) is based on the momentum conservation law, while equations (1b) and (1c) are related to the energy conservation law for the total energy density ρ . These laws can be combined into the conservation of the energy momentum tensor $T_{\mu\nu}$ (ref. 14, equation 126.1), which in arbitrary coordinates reads

$$T^{\mu\nu}_{;\nu} = 0 \quad (3)$$

Let us now consider Einstein's equations,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R; \quad G_{\mu\nu} = -8\pi GT_{\mu\nu}; \quad (4a)$$

$$G^{\mu\nu}_{;\nu} = 0 \quad (4b)$$

Equation (4b) is the Bianchi identity, a geometrical relation valid irrespective of the content of the right-hand side of equation (4a)¹⁵. Clearly equations (1), or the global form, equation (3), are compatible with equation (4b) only if G is constant.

This argument indicates that taking G in equation (1a) variable is inconsistent with the simultaneous use of equations (1b) and (1c) in which G is absent not because it has no *a priori* right to be there, but because it has already been assumed to be constant.

The Einstein framework

Conventional general relativity is based on the existence of a metric field which obeys the Einstein field equations (4b) with or without the right-hand side (ref. 16 and P. Bergmann, personal communication), $GT_{\mu\nu}$, a quantity that Einstein¹⁷ considered "a formal condensation of all those things whose comprehension in the sense of a field theory is still problematic". Equations (4a) were constructed from left to right¹⁵. Equation (4b) means that the right-hand side of equations (4a) must have zero divergence. While one may choose G to be constant, thus yielding $T^{\mu\nu}_{;\nu} = 0$ which embodies the 'standard' equations of motion and the 'standard' energy conservation equation, this choice is clearly not dictated by the theory of gravity. For that reason Einstein considered his theory 'incomplete', in the sense that it did not prescribe how to construct the source term from within. Given equation (4b), it would have been equally possible to choose

$$(GT^{\mu\nu})_{;\nu} = 0 \quad (5)$$

with G as a variable. Equation (5) indicates that $T^{\mu\nu}_{;\nu} = 0$, equation (3), is no longer true, because there is now a source term proportional to $G_{;\nu}$. Because the energy conservation equation has acquired new terms involving G , so will the resulting equation of state. To summarize, as the energy-momentum tensor $T_{\mu\nu}$ is a function of the pressure p and the total energy density ρ , we have from equation (4) the two possibilities:

G constant plus Bianchi identity $\rightarrow T^{\mu\nu}_{;\nu} = 0 \rightarrow p = p(\rho_0)$, $M = \text{constant}$

G variable plus Bianchi identity $\rightarrow (GT^{\mu\nu})_{;\nu} = 0 \rightarrow p = p(\rho_0, G)$, $M \neq \text{constant}$

independently of the specific form of $T_{\mu\nu}(p, \rho)$. To illustrate the modification to the p versus ρ relation, consider the following form for $T_{\mu\nu}$

$$T_{\mu\nu}(p, \rho) = (p + \rho)u_\mu u_\nu - pg_{\mu\nu} \quad (6)$$

By introducing the variables $p_* = pG$ and $\rho_* = \rho G$, we formally

reduce equation (5) to the standard form $T^{\mu\nu}_{;\nu} = 0$. Because G has 'disappeared', we recover again, as in the standard case $p_* \sim (\rho_0)_*$ and $M_* = \text{constant}$. Returning to the variables p and ρ_0 , we obtain finally

$$p \propto \rho_0^\gamma G^{\gamma-1} \quad (7a)$$

$$MG = \text{constant} \quad (7b)$$

Alternatively, as $T^{\mu\nu}_{;\nu} = 0$ implies $p_* dV + dU_* = 0$, on integrating, $(pV = NkT = \Gamma U)$, $NTGV^\Gamma = \text{constant}$. However, the zero pressure limit of $T^{\mu\nu}_{;\nu} = 0$ implies $nVG = NG = \text{constant}$, so that substituting T in $pV = NkT$, we recover equation (7a) (see also ref. 4).

As for the hydrostatic equilibrium relation, the introduction of the new variables p_* and ρ_* has no effect on its structure, because G depends only on time. Equation (1a) is therefore consistent with a varying G .

Inserting equation (7a) into equation (1a), we derive,

$$R \sim M^{(\gamma-2)/(3\gamma-4)} G^{-1/(3\gamma-4)} G^{(\gamma-1)/(3\gamma-4)} \\ \sim (GM)^{(\gamma-2)/(3\gamma-4)} \sim \text{constant} \quad (8)$$

a result radically different from equation (2).

Alternatively, we can say that because only the starred quantities are physically meaningful, equations (1) should read ($\rho_0 = \rho$ for ease of notation)

$$\frac{1}{\rho_*} \frac{dp_*}{dr} = -\frac{m_*}{r^2}, \quad p_* \propto \rho_*^\gamma, \quad M_* = \text{constant} \quad (9)$$

which clearly show that G does not appear.

Scale covariant framework

The previous analysis has indicated that while it is in principle possible formally to make room for a variable G within the Einstein framework, the results turn out to be identical to those with a constant G . The message is that the quantity G has no 'identity' in such framework, thus confirming that there is no useful way of talking about a variable G within the Einstein theory.

To incorporate fully the possibility that atomic and gravitational times may differ over large fractions of the age of the Universe a formalism, the scale covariant theory (SCT) of gravity, was devised. A gauge function $\beta(t)$, epitomizing our ignorance of how gravity and atomic forces couple, was introduced through the equation

$$\Delta t_E = \beta(t) \Delta t \quad (10)$$

While the strong equivalence principle assumes $\beta(t) = \text{constant}$, a broad-based analysis of observational data ranging from astrophysics to cosmology (up to $z \approx 10^3$), gave no indication that $\beta(t)$ has to be constant^{3,4,6,7,18}. While this does not mean that $\beta(t)$ must be variable, it is a significant result once we consider that it was firmly believed that even simple data like the luminosity of the Sun would have clearly opposed the variability of $\beta(t)$. Furthermore, recent measurements of the period of the Moon by both atomic and gravitational clocks do not seem to yield the same result, the residual corresponding to

$$\frac{\dot{\beta}}{\beta} = (2.75 \pm 0.64) 10^{-11} \text{ yr}^{-1} \quad (11)$$

Gravitational times are defined as those times with respect to which the Einstein equations retain exactly their form as do all the purely gravitational expressions ensuing therefrom. However, with respect to atomic units, the gravitational equations take the form (ref. 6, equation (2.10))

$$G_{\mu\nu} + f_{\mu\nu}(\beta) = -8\pi GT_{\mu\nu}(\beta) \quad (12)$$

The final aim of the SCT is to construct a viable two-times theory, within which the commonly assumed identity between atomic and gravitational times may be broken, $\Delta t \neq \Delta t_E$. For that to occur, the lagrangian $\mathcal{L}$ describing gravity and matter must not be scale invariant. While it is not *a priori* clear which of the

terms in $\mathcal{L}$ should break the symmetry, the choice is not large. Brands and Dicke (BD) chose the gravitational part $\mathcal{L}_g$, leaving $\mathcal{L}_m$, the matter part, unaltered. This implies that their $T_{\mu\nu}$ does not depend on β : the conservation law (3) holds unchanged, as do relations (1). Relation (2) is therefore consistent with the BD framework. However, the change in $\mathcal{L}_g$ means that the Einstein equations are modified by the addition of terms which coincide with our function $f_{\mu\nu}(\beta)$ only if the BD parameter ω is $-3/2$. The potentially most serious source of troubles facing a theory that alters the Einstein equations is that the expressions for the three fundamental tests acquire extra contributions which may spoil the ever increasing agreement with observations. As this seems to be the case in the BD theory, the SCT was constructed with the opposite point of view. The scale breaking terms were relegated entirely to $\mathcal{L}_m$, while the gravitational part of $\mathcal{L}$ was constructed in a scale invariant manner. The result is equation (12), where the function $f_{\mu\nu}(\beta)$ assures that the left-hand side of equation (12) is scale invariant, that is it has power zero under the scale transformation $ds \rightarrow ds' = \beta(x) ds$, where for any quantity A , we have $A \rightarrow A' = \beta^{\pi(A)} A$, $\pi(A)$ being the 'power' of A .

This approach has been questioned¹⁹ on the grounds that the 'presently accepted microscopic lagrangian' can be shown to be scale invariant (a more detailed discussion is given elsewhere). One of the critical quantities entering the analysis of ref. 19 is $\pi(m)$, the power of a microscopic mass m . From the fact that $\hbar/mc$ is a length, it follows that $\pi(l) = \pi(\hbar) - \pi(m) = +1$. The claim¹⁹ that $\pi(m) = -1$ can only be made if $\pi(\hbar) = 0$, which in turn implies $\pi(e) = 0$, because $e^2/\hbar c$ is a pure number. Now because the electric charge enters only in the lagrangian term representing interactions, any assumption about $\pi(e)$ is an assumption about the scaling property of the interaction. Because of the general relation $\beta GM = \text{constant}$ (see equation (14)), $\pi(m) = -1$ implies $g = 2$. Now considering the electromagnetic tensor $F^{\mu\nu}$, if the contribution to $T_{\mu\nu}$ from the free electromagnetic field is taken to be scale invariant then the power of $F^{\mu\nu}$ is $-3 - g/2$ which is -4 . Because $J^\mu = enu^\mu$ has power -4 , the interaction equation $F^{\mu\nu}{}_{;\nu} = J^\mu$ is scale invariant. As such, it may not hold in the SCT because the final aim is that of allowing atomic clocks to scale differently from gravitational clocks. Because the gravitational part $\mathcal{L}_g$ has been constructed in a scale invariant form (hence scale invariant theory of gravitation) with the result that the period of a planet with respect to atomic clocks scales like $P \sim \beta^{-1}$ (ref. 6, equation (4.19)), clearly the dynamic equation governing an atom cannot be $F^{\mu\nu}{}_{;\nu} = J^\mu$ because being scale invariant, it would yield $P(\text{atom}) \sim \beta^{-1}$ too, thus leading to no difference between atomic and planetary dynamics, that is to a constant G . If we accept "the presently accepted form of $\mathcal{L}_m$ "¹⁹ (an equation like $F^{\mu\nu}{}_{;\nu} = J^\mu$) then clearly the only possibility for a varying G is to go back to $\mathcal{L}_g$ and to devise a modified BD approach. This line of reasoning, which does not contemplate the possibility advocated by the proponents of the SCT, cannot be taken to imply nor to have proved that either (1) G is constant or (2) that for it to vary, the only avenue open is $\mathcal{L}_g$. Progress in modifying $F^{\mu\nu}{}_{;\nu} = J^\mu$ using gauge fields and neutral current concepts, has recently been made (P.J. Adams and J. Anderson, personal communication). Far from being mutually exclusive, the approach proposed in ref. 19 and the point of view of the SCT are in our opinion but different approaches towards the same goal.

Let us now prove that βGM is constant. It has been shown⁶ that the left-hand side of equation (12) has zero co-covariant derivative (indicated by * instead of ;). We therefore have, using equation (A.19) of ref. 6 and the fact that $G_* = 0$,

$$T^{\mu\nu}{}_{;\nu} + (\pi + 6)\phi_\nu T^{\mu\nu} - \phi^\mu T^\nu{}_\nu = 0 \quad (13)$$

with $\phi = \ln \beta$, $\phi_\nu = \phi_{,\nu}$. The quantity π is the power of $T^{\mu\nu}$. Multiplying equation (13) by u_μ and using for $T_{\mu\nu} = \rho_0 u_\mu u_\nu$, we obtain on integrating ($\rho_0 V = M$)

$$\beta GM = \text{constant} \quad (14)$$

where $V/V = u^\mu{}_{;\mu}$, and $\pi = \pi_{S1} = -4 - g$, $g = \pi(G)$. We have

required that the right-hand side of equation (12) be scale invariant, $\pi(GT_{\mu\nu}) = 0$. Equation (14) makes it clear that the power of m must be chosen consistently with the dynamical constraints of the theory, and not assumed *a priori*. In fact because $\pi(m) = 1 - g$, a $G\beta^2 = 1$ gauge yields $\pi(m) = -1$, whereas $G\beta = 1$, the so-called non-matter creation case, yields $\pi(m) = 0$.

The very fact that in equation (12) $T_{\mu\nu}$ is taken to depend on β implies, even without further qualifications, that we cannot use, without checking their validity, any of the standard relations (1a), (1b) and (1c) because, strictly speaking, they are valid only for constant β . The β dependence of microscopic relations, marking a sharp contrast with the BD theory, is the price one must pay for constructing a theory with a built in two-times scheme and which at the same time preserves the agreement between observations and predictions for the fundamental tests, whose expressions in atomic units can be derived by direct scaling of the corresponding gravitational expressions. Because at any given time the scale function $\beta(t)$ can be normalized to unity, a set of measurements at one time cannot reveal the presence of β . In this sense, the SCT theory has preserved agreement with the three tests⁷. The scaling of gravitational expressions to obtain the corresponding atomic ones is all that is required when dealing with purely gravitational phenomena, like periods or distances of planets, which by assumption are given by Einstein equations and are constant with respect to gravitational clocks. On the other hand, non-gravitational quantities must be derived consistently with equation (13), so that the influence of the function β can be fully accounted for. This is a major endeavour because it calls for a reconstruction of microscopic physics, a process which, while accounting for the fact that β can enter at any stage, must satisfy the following two constraints: (1) the atomic clock ensuing from such theory must not depend on β ; (2) the many-body behaviour of the theory must yield results consistent with the classical ones derived from equation (13). A typical example of such 'matching' relation is the continuity equation, $\dot{\rho} + \rho \text{div } j = 0$, which can be obtained both from equation (3) and from the Schrödinger equation. These requirements clearly indicate that the attempt at constructing such theory is far more complex and difficult than that faced by Brans and Dicke. Descending from the Einstein equation to microphysics (thus running in the opposite direction to the traditional one, which assumes the validity of the microscopic scheme and makes connection with gravity by going from a flat to curved space) has to my knowledge never been worked out in detail. In spite of this, we have already derived⁶ the influence of β on several microscopic relations, and the corresponding observational implications such as the energy versus frequency for a free photon; the adiabatic scaling law for a perfect gas; the equilibrium radiation versus frequency and temperature expression; and the luminosity versus distance relation (see also ref. 5). No contradictions have yet appeared.

Having clarified this point, let us return to the original problem of evaluating R as $R(t)$. The full problem is very complex because we do not yet possess a full description of atomic interactions which determine the equation of state for an object like the Earth. We can therefore present only a hopefully good approximation to the exact solution. Returning to equation (13), and using for $T_{\mu\nu}$ equation (6), we first separate out the rest mass contribution from the total energy density ρ , $\rho = \rho_0 + u$. The rest mass part of equation (13) can be easily integrated. The result is equation (14). The $\rho - \rho_0 = u$ part satisfies the equation ($U = uV$)

$$dU + p dV + [(1 - g)U + 3pV] d\phi = 0 \quad (15)$$

which is the first law of thermodynamics generalized to include the β terms³. Let us now write

$$p = \frac{N}{V} kT \left[1 + B(T) \frac{N}{V} + C(T) \left(\frac{N}{V} \right)^2 + \dots \right] \quad (16)$$

where $B(T)$ and $C(T)$ are the so-called virial coefficients²⁰. We first consider the perfect gas case, $B = C = 0$. Using $pV = NkT$

and $U = NkT/\Gamma$ and remembering that $N \sim (\beta G)^{-1}$, we integrate equation (15), with the results ($\gamma = 1 + \Gamma$)

$$p \propto \rho_0^\gamma \left(\frac{G}{\beta^2} \right)^{\gamma-1} \quad (17a)$$

$$T\beta^{3\Gamma} \sim \frac{1}{V^\Gamma} \quad (17b)$$

For a more general derivation, valid for non-adiabatic transformations, see ref. 5.

Note that there is no choice of gauge, a G versus β relation, that reduces equations (17a) and (14) to equations (1b) and (1c), at least for $\gamma \neq 1$. In fact, to recover equation (1b), we must choose $G \sim \beta^2$, which implies $\dot{\beta} < 0$, contrary to equation (11). This choice would also imply $M \sim \beta^{-3}$, not $M \sim \text{constant}$. To see the consequences of equation (17a), we use it in equation (1a), which has already been shown⁶ to be valid in the present framework:

$$R \sim M^{(\gamma-2)/(3\gamma-4)} G^{-1/(3\gamma-4)} \left(\frac{G}{\beta^2} \right)^{(\gamma-1)/(3\gamma-4)} \sim \frac{1}{\beta} \quad (18)$$

on using equation (14). The result is independent of γ . We therefore conclude that

$$\frac{\dot{R}}{R} = -\frac{\dot{\beta}}{\beta} = -(2.75 \pm 0.64) 10^{-11} \text{ yr}^{-1} \quad (19)$$

where we have used equation (11). The predicted value of R , 400 Myr ago, is therefore ($R_0 = 1$)

$$R = 1.0110 \pm 0.002 \quad (20)$$

which can be compared with the recent palaeomagnetic result¹²

$$R = 1.020 \pm 0.028 \quad (21)$$

At first sight, the close agreement can be taken as justification for the approximations made to arrive at equation (18). However, a closer scrutiny reveals that the result must be taken with great care. First, as stressed in ref. 12, an equation of state of the form of equation (17a), is not valid at the surface of the planet where $p = 0$ but $\rho_s \neq 0$. The criticism is clearly independent of G and β . More important, however, is that the result $R \sim \beta^{-1}$ implies that the radius of the Earth in einstein units $R_E = \beta R$ is constant. While possible, this result is not obvious because atomic quantities like e and m , which enter the relation p_E as $p_E(\rho_E)$ and therefore R_E , are not constant in einstein units.

Having discussed the limitations of equation (17a), we go back to equation (16) and include the interaction term $B(T)$ in the form $B(T) = -BT^{\gamma-1}$. We simulate in this phenomenological way an attractive potential. Although equation (17b) is valid when $B = 0$, it is used because we are performing an approximation to the first term. Equation (16) then becomes ($\rho_0 = \rho$ for ease of notation)

$$p = A\rho^\gamma \left(\frac{G}{\beta^2} \right)^{\gamma-1} \left[1 - \frac{B}{A} \rho^{\tilde{\gamma}} \left(\frac{G}{\beta^2} \right)^{\tilde{\gamma}-1} \right] \quad (22)$$

where $\gamma = 1 + \Gamma$, $\tilde{\gamma} = \gamma_* - \gamma$, $\gamma_* = 2 + (\gamma - 1)b$. An equation of state for the Earth of the type (22) (for $G = 1$ and $\beta = 1$) has been proposed by Birch²¹ with $\gamma = 7/3$ and $\gamma_* = 5/3$. While equation (17a) allows an exact solution, equation (18), equation (22) does not. For our purposes

$$R \sim \frac{1}{\beta} \left[1 - \frac{B}{A} \bar{\rho}^{\tilde{\gamma}} \left(\frac{G}{\beta^2} \right)^{\tilde{\gamma}-1} \right]^{1/(3\gamma-4)} \quad (23)$$

where $\bar{\rho}$ is an average (but time dependent) density. We further derive

$$\frac{\dot{R}}{R} = -r \frac{\dot{\beta}}{\beta}; \quad r - 1 = \frac{(2+g)\alpha}{\mathcal{D} - 3\tilde{\gamma}\alpha} \quad (24)$$

where $\mathcal{D} = 3\gamma - 4$, $\alpha = x(1-x)^{-1}$, $x = (B/A)\bar{\rho}^{\tilde{\gamma}} (G/\beta^2)^{\tilde{\gamma}-1}$. For the values of γ and γ_* given by Birch, we have $r - 1 = (2+g)\alpha/(3+2\alpha)$, which for $g = 1$ ranges from zero ($\alpha = 0$) to $3/2$ ($\alpha =$

∞). The first value corresponds to equation (18); $\alpha = \infty$ implies $x = 1$, that is $\bar{\rho} = \rho_s$, not acceptable. If p must vanish at $\rho = \rho_s$, then $x = (\bar{\rho}/\rho_s)^{\tilde{\gamma}}$. In spite of the large excursion in α , the corresponding variation in $r - 1$ is only from 1 to 1.5. Taking $\alpha = 1$ ($x = 1/2$, corresponding to $\bar{\rho} = 8 \text{ g cm}^{-3}$, $\rho_s = 2.84 \text{ g cm}^{-3}$) and $g = 1$, we have $r = 1.6$, so that 400 Myr ago

$$R = 1.018 \pm 0.004 \quad (25)$$

in better agreement with equation (21). Finally equation (24) may alternatively be written, using equation (14),

$$R \sim (GM)^r, \quad \frac{\dot{R}}{R} = r \frac{(\dot{GM})}{GM} \quad (26)$$

to be compared with equation (2). As r is positive (at least for $g + 2 > 0$) the SCT predicts a result qualitatively different from equation (2): for a constant mass M , a larger G implies a larger radius, not a smaller one.

The method used to arrive at equation (22) is not meant to be a microscopic derivation of the p versus (ρ, β) relation for the Earth, but only an heuristic approach to show how β may enter the problem. Even without β , it is difficult to derive²² an equation of state for the Earth. Because all the $p = p(\rho)$ relations discussed in ref. 22 contain no β and because $\beta = 1$ corresponds to einstein units, one might be tempted to think that the relations in ref. 22 are actually valid in gravitational units and that a simple scaling could then yield the corresponding atomic relations. This is, however, not the case. Such a procedure would yield a relation where the difference of the power of ρ and that of (G/β^2) is unity, like the first term of equation (22). In turn this yields $r = 1$: it can be seen that $r - 1$ is directly proportional to the difference of these two exponents. Furthermore, $r = 1$ implies $R_E = \text{constant}$ producing difficulties already discussed. With the present procedure, this does not happen. The powers of ρ and that of (G/β^2) of the second term of equation (22) do not differ by one but by two. An extra term $(G/\beta^2)^{-1}$ appears which makes $r \neq 1$. This extra factor, which breaks the symmetry that one would invariably get out of a simple scaling process, cannot be arrived at from gravitational units where $\beta = \text{constant}$.

Lunar data

Having derived equation (24) but having no estimates for $\dot{\beta}/\beta$ from within the theory, we can either use palaeomagnetic data for $\dot{R}/R$ and thus determine $\dot{\beta}/\beta$ (and possibly $\dot{G}/G$), or we evaluate $\dot{\beta}/\beta$ from other data and then compare the resulting $\dot{R}/R$ with palaeomagnetic data. The only other data available are the rate of change of the Moon's period measured in both atomic and gravitational units: these are ($P = 2\pi/n$ is the period of revolution of the Moon)

$\dot{n}_E$ (gravitational time) (arc s cyr ⁻²)		$\dot{n}$ (atomic time) (arc s cyr ⁻²)	
-26.0 ± 2.0	ref. 23	-21.4 ± 2.6	ref. 26
-28.5 ± 3.1	ref. 23	-23.6 ± 1.5	refs 26, 27
-30.0 ± 3.0	ref. 7	-24.6 ± 3.9	refs 26, 28
-27.4 ± 3.0	ref. 24		
-30.6 ± 3.1	ref. 25		

where, because of equation (10), $n = \beta n_E$. A least-square fit analysis yields ($\dot{\beta}/\beta$ in 10^{-11} yr^{-1}); note that $1 \text{ cyr} = 10^2 \text{ yr}$

$$\text{All equations} \quad \dot{n}_E = -27.97 \pm 0.78; \quad \dot{\beta}/\beta = 2.75 \pm 0.64 \quad (28a)$$

$$\text{First } \dot{n}_E \text{ deleted} \quad \dot{n}_E = -29.11 \pm 0.73; \quad \dot{\beta}/\beta = 3.41 \pm 0.54 \quad (28b)$$

$$\text{Second } \dot{n}_E \text{ deleted} \quad \dot{n}_E = -27.87 \pm 0.91; \quad \dot{\beta}/\beta = 2.69 \pm 0.72 \quad (28c)$$

Because the value of $\dot{\beta}/\beta$ retaining all the eight equations turns out to fall between the other two, we have adopted this value, equation (11). Finally, independently of the way we determine $\dot{\beta}/\beta$ (palaeomagnetic or lunar data), the value of $\dot{G}/G$ is gauge dependent and is to be determined using the relations $\dot{G}/G = -g\dot{\beta}/\beta$, $\pi(m) = 1 - g$.

Moment of inertia

Using equation (24), and $M = \text{constant}$ we obtain $\dot{I}/I = -2r\dot{\beta}/\beta$, or equivalently $\dot{I}/I = 2r\dot{G}/G$, with $r > 0$. This result agrees with equation (A.16) of ref. 29, because the G in equation (A.1) must be written as $G_* = G(\beta^2/G)^{\gamma-1} = G^{4-3\gamma}$, due to equation (17a). If so, equation (A.16) becomes $\dot{I}/I = a\dot{G}/G$, with $a = (3\gamma - 4)/10 > 0$.

This result is, however, of limited use because it is in atomic units and it represents only the cosmological and not the total contribution to $\dot{I}/I$, whereas the estimates reported in the literature are total variations—due to all possible causes, cosmological and otherwise—and they are all with respect to gravitational units. To prove this last point, we must remember that the basic ingredient (see refs. 30, 31) is the conservation of the total angular momentum for the Earth–Moon system $J_E + L_E = \text{constant}$. If the Earth is considered a point mass, this implies the conservation of the orbital angular momentum L_E , a law valid only with respect to gravitational units^{6,7}. In atomic units L scales like $L \sim \beta^2/G$ ($\sim G^{-3}$ for constant mass) (ref. 6, equation 4.7).

We now consider $\dot{I}_E/I_E$ from all sources and expressed in gravitational units. While growth lines in fossils promised to be a valuable source of information³⁰, recent difficulties in interpretation have indicated that quantitative results are not yet possible³². We shall therefore rely on ancient astronomical data (see Chapter 10, ref. 31), and take equation (10.2.2b) from ref. 31 and generalize it to include a non-constant I_E . We obtain

$$\frac{\dot{I}_E}{I_E} = -\frac{\dot{\Omega}_E}{\Omega_E} + \frac{\gamma}{3} \frac{\dot{n}_E}{n_E} + H \quad (29)$$

where $\gamma = (1+b)L_0/J_0$, b accounting for the solar contribution. H represents the contribution of the time variation of dynamical elements describing the size, shape and orientation of the Earth–Moon system. If $J = I\Omega \cos \phi$, and $L = M_e M_m / (M_e + M_m) \times R^2 n (1-e^2)^{1/2} \cos i$, then $H = H(\dot{e}, \dot{\phi}, di/dt)$ (see ref. 33).

Ancient astronomical observations can be used³¹ to construct a relation of the form $X_i \dot{n}_E + Y_i \dot{\Omega}_E = f$, where X_i and Y_i are numerical coefficients and f is a function of the observed less the computed position of the recorded event. As stated by Lambeck³¹ “the computed positions and locations” are arrived at “assuming purely gravitational motion”, further substantiating our claim that the ingredients are the standard newtonian equations which hold only with respect to dynamical units with G_E and M_E constant. Lambeck summarizes the results as follows

$$\dot{n}_E = -28 \pm 2; \quad \dot{\Omega}_E = -1,100 \pm 100 \quad (30)$$

This value for $\dot{n}_E$ is not substantially different from the one determined in equation (28). Using $L_0/J_0 = 4.89$ and $b = 0.21$, we obtain from equation (29)

$$\frac{\dot{I}_E}{I_E} = -(0.086 \pm 0.031) 10^{-9} \text{ yr}^{-1} + H \quad (31)$$

To complete the computation we must now evaluate and then subtract from equation (31) the purely cosmological contribution. This is done by remembering that $I_E \sim M_E R_E^2 \sim \beta^2 R^2$, and using equation (24)

$$\left(\frac{\dot{I}_E}{I_E}\right)_{\text{cosm}} = 2 \frac{(\dot{\beta}R)}{\beta R} = -2(r-1) \frac{\dot{\beta}}{\beta} = -(0.033 \pm 0.008) 10^{-9} \text{ yr}^{-1} \quad (32)$$

The difference

$$\frac{\dot{I}_E}{I_E} - \left(\frac{\dot{I}_E}{I_E}\right)_{\text{cosm}} = -(0.053 \pm 0.032) 10^{-9} \text{ yr}^{-1} + H \quad (33)$$

may now be attributed to geophysical phenomena such as the formation of the Earth's core³⁰. For the evaluation of H see ref. 33.

Conclusions

The question whether a viable theory of gravitation exists that permits equations (1a), (1b) and (1c) as they are, while G is taken to vary, has been investigated with the following results.

Standard Einstein theory can be formally generalized to include a non-constant G . While equation (1a) remains valid, both equations (1b) and (1c) change in such a way that the final result is $R = \text{constant}$, demonstrating that G is actually a non-entropy within the strict framework of Einstein.

In the BD theory, as by construction the matter lagrangian is not changed, equations (1) and (2) both hold. However, it is believed that the BD theory faces difficulties, at least in its present form.

Finally, while equation (1a) holds unchanged in the SCT equations (1b) and (1c) cannot be simultaneously satisfied. The most interesting result, however, which transcends the particular example equation (22), is that microphysics may be affected by a variable G . This may seem unusual at first as microphysics contains no G thus creating the impression that it ought to be immune from whatever happens to G . However, relations like the Boltzman equation, the radiative transfer equation, and the Navier–Stokes equations are derivable from and are consistent with the conservation law $T^{\mu\nu}_{;\nu} = 0$ for the energy momentum tensor, a law which holds only if G is constant, as equation (13) shows. The absence of G is actually a consequence of having chosen it to be constant.

More general derivations of classical non-gravitational relations (such as, Liouville equation⁵, Boltzman distribution⁵, radiative transfer equation⁵ and classical thermodynamics^{3,6}) in which β does explicitly appear, are possible, if constructed consistently with the general conservation laws that follow from the Einstein equation. For example, an equation of state of the form (17a) with $\gamma = 4/3$ holds^{3,6} for radiation (ρ_0 is replaced by the photon number density n_γ). Because $3p_\gamma = \rho_\gamma$ and $n_\gamma T \sim \rho_\gamma$, $\rho_\gamma \sim (\beta^2 G^{-1}) T^4$. It then follows that the Sun's luminosity $L_\odot$ scales like $L_\odot \sim \beta^{-1}/k$, where k is the opacity. For constant k and constant rest mass, $L_\odot \sim G$, contrary to the early statements that $L_\odot \sim G^7$ (for details, see refs 4, 5).

Our comparisons of the theoretical predictions with observations must be treated cautiously. The value of $\dot{\beta}/\beta$ derived from lunar data and used to estimate equations (20) and (25) is still subject to uncertainties due to the difficulties in estimating the errors; there are also doubts concerning palaeomagnetic data, or more precisely the assumptions made to arrive at equation (21) (ref. 10; see, however, ref. 34). Finally the evaluation of the function H , equation (33), is still fraught with large errors (see ref. 33).

We prefer, therefore, to focus the dependence of R on G . If we write, in atomic units,

$$\frac{\dot{R}}{R} = \left(\frac{\dot{R}}{R}\right)_{\text{cosm}} + \left(\frac{\dot{R}}{R}\right)_{\text{othersources}} \quad (34)$$

the present work predicts a reversal (compared with the standard treatment, equation (2)), of the sign of the first term in equation (34) when written in terms of $\dot{G}/G$, equation (26).

This is in turn due to the new effect represented by the presence of G in the $p = p(\rho)$ relation: the G in the right-hand side of equation (1a) is overcompensated by the $G^{\gamma-1}$ factor in equation (17a), thus causing a larger G to correspond to a larger radius. However, the evaluation of the second term in equation (34) is outside the scope of this article, and we cannot conclude from our work alone that R is actually decreasing.

I thank Drs G. Blake, D. J. Stevenson and J. Anderson for constructive criticism, also Drs T. C. Van Flandern, S.-H. Hsieh, P. J. Adams, I. Goldman, J. R. Owen, K. M. Creer and J. D. Mulholland for assistance during the preparation of this manuscript.

Received 8 July 1980; accepted 9 February 1981.

1. Will, C. M. in *General Relativity* (eds Hawking, S. W. & Israel, W.) (Cambridge University Press, New York, 1979).
2. Dirac, P. A. M. *Nature* **139**, 323 (1937).
3. Canuto, V. M., Adams, P. J., Hsieh, S.-H. & Tsiang, E. *Astrophys. J. Suppl.* **41**, 243 (1979).
4. Canuto, V. M. & Hsieh, S.-H. *Astrophys. J.* **237**, 613 (1980).
5. Canuto, V. M. & Hsieh, S.-H. *Astrophys. J.* (in the press).
6. Canuto, V. M., Adams, P. J., Hsieh, S.-H. & Tsiang, E. *Phys. Rev. D* **16**, 1643 (1977) (Paper I); **41**, 263 (1979) (Paper II); **41**, 301 (1979).
7. Canuto, V. M., Hsieh, S.-H. & Owen, R. J. *Mon. Not. R. astr. Soc.* **188**, 829 (1979).
8. Teller, E. *Phys. Rev.* **73**, 801 (1948).
9. Jordan, P. *The Expanding Earth* (Pergamon, New York, 1971).
10. Carey, S. W. *The Expanding Earth* (Elsevier, Amsterdam, 1976).
11. Wesson, P. S. *Cosmology and Geophysics* (Adam Hilger, Bristol, 1978).
12. McElhinny, M. W., Taylor, S. R. & Stevenson, D. J. *Nature* **271**, 316 (1978).
13. Bishop, W. T. & Landsberg, P. T. *Nature* **264**, 346 (1976).
14. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* (Addison-Wesley, Reading, 1959).
15. Dirac, P. A. M. *General Theory of Relativity*, 25–26; 43–45 (Wiley, New York, 1975).
16. Bergmann, P. *Introduction to the Theory of Relativity* (Prentice Hall, Englewood Cliffs, 1960).

17. Einstein, A. in *Autobiographical Notes* (ed. Schilpp, P. A.) 71, (Open Court, Chicago, 1979).
18. Canuto, V. M. & Hsieh, S.-H. *Phys. Rev. Lett.* **44**, 695 (1980).
19. Bekenstein, J. D. & Meisels, A. *Phys. Rev. D* **22**, 1313 (1980).
20. Landau, L. D. & Lifshitz, E. M. *Statistical Physics*, 224 (Pergamon, New York, 1958).
21. Birch, F. J. *geophys. Res.* **57**, 227 (1952).
22. Stacey, F. D., Brennan, B. J. & Irvine, R. D. *Geophys. Surv.* (in the press).
23. Morrison, L. V. & Ward, C. G. *Mon. Not. R. astr. Soc.* **173**, 183 (1975).
24. Goad, C. C. & Douglas, B. C. *J. geophys. Res.* **83**, 2306 (1978).
25. Lambeck, K. *Phil. Trans R. Soc.* **287**, 545 (1977).
26. Van Flandern, T. C. *Astrophys. J.* (in the press).
27. Dickey, J. O., Williams, J. G. & Yoder, C. F. *EOS* **61**, 46, 939 (1980).
28. Calame, D. & Mulholland, J. D. *Science* **199**, 977 (1978).
29. Nordvedt, K. & Will, C. M. *Astrophys. J.* **177**, 755 (1972).
30. Runcorn, S. K. *Nature* **193**, 311 (1962); **195**, 1248 (1962); **204**, 823 (1964).
31. Lambeck, K. *The Earth's Variable Rotation: Geophysical Causes and Consequences* (Cambridge University Press, 1980).
32. Scrutton, C. T. in *Tidal Friction and the Earth's Rotation* (ed. Broshe, P.) (Springer, New York, 1978).
33. Hipkin, R. G. in *Growth Rhythms and the History of the Earth's Rotation* (eds. Rosenberg, G. D. & Runcorn, S. K.) (Wiley, New York, 1975).
34. Schmidt, P. W. & Clark, D. A. *Geophys. J. R. astr. Soc.* **61**, 95 (1980).

Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA

David B. McKay & Thomas A. Steitz

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, USA

The 2.9 Å resolution crystal structure of Escherichia coli catabolite gene activator protein (CAP) complexed with cyclic AMP reveals two distinct structural domains separated by a cleft. The smaller carboxy-terminal domain is presumed to bind DNA while the amino-terminal domain is seen to bind cyclic AMP. Model building studies suggest that CAP binds to left-handed B-type DNA, contacting its major groove via two α -helices. It is possible that the CAP conversion of right- to left-handed DNA in a closed supercoil, is what activates transcription by RNA polymerase.

REPRESSOR and activator proteins that regulate gene expression at the level of transcription recognize specific nucleotide sequences in double-stranded DNA. Although it has been suggested that specific recognition involves α -helices fitting into the major groove of B-DNA^{1–3} or anti-parallel β -strands in the minor groove⁴, no experimental evidence exists concerning the actual mode of sequence-specific interaction between protein and double-stranded DNA. Also, the molecular mechanism by which gene activators promote the activity of RNA polymerase, thereby switching on genes, remains unknown. We present here the structure determination of a protein that binds in a sequence-specific manner to double-stranded DNA and propose a mechanism by which the catabolite gene activator protein is able to switch on the catabolite-sensitive genes.

The catabolite gene activator protein (CAP), also called the cyclic AMP receptor protein, functions in *Escherichia coli* in the regulation of several catabolite-sensitive gene operons^{5,6}. Regulation by CAP is exerted at the transcriptional level with cyclic AMP acting as an allosteric effector. In the presence of a sufficient concentration of intracellular cyclic AMP, cyclic AMP forms a complex with CAP which binds to specific DNA sites near the promoters of several operons and alters the rate of their transcription by RNA polymerase. In the lactose (*lac*)⁷ and arabinose⁸ operons, the cyclic AMP–CAP complex is a positive regulator—it potentiates transcription. In the galactose operon, the presence of two overlapping promoters for RNA polymerase makes the situation more complex; cyclic AMP–CAP is required for initiation of transcription from one promoter, but inhibits transcription from the other. Thus, in this case it apparently acts as both a positive and negative regulator of operon expression⁹.

The active form of CAP is a dimer of identical subunits of molecular weight 22,500 and 201 amino acid residues^{10,11}. Proteolytic cleavage studies are consistent with the CAP subunit

having two separate structural domains¹². Results from a variety of techniques^{12–17} suggest that CAP undergoes a significant conformational change on binding cyclic AMP.

To develop a structural basis for understanding (1) the cyclic AMP-induced allosteric transition in CAP, (2) the site-specific recognition of DNA by CAP, and (3) the mechanism of transcription activation by CAP, we have initiated crystallographic studies of CAP and its complexes with ligands. In this article we report the structure of the cyclic AMP–CAP complex at 2.9 Å resolution.

Structure determination

CAP protein was purified and crystallized in the presence of 0.5 mM cyclic AMP as previously described¹⁸. The crystals are orthorhombic, space group P2₁2₁2₁, $a = 46.5$ Å, $b = 97.1$ Å, $c = 105.4$ Å, with one dimeric CAP molecule per asymmetric unit.

Intensities of crystallographic reflections were measured by diffractometer, using the Wyckoff scan algorithm¹⁹. Friedel pairs of reflections were measured on native crystals and five heavy-atom derivatives; for derivative data sets only ~75% of the total reflections, consisting of those which were most intense in the native data, were measured. Heavy-atom positions were located and refined by conventional methods²⁰ and the correct enantiomorph of the heavy-atom positions was determined using anomalous difference Fourier²¹. A summary of heavy-atom derivative preparation and refinement statistics is presented in Table 1. Combined multiple isomorphous replacement and anomalous scattering phases were computed to 2.9 Å resolution and had an average figure of merit of 0.74.

Part of the electron density map is displayed in Fig. 1. Most of the polypeptide backbone is well ordered and can be traced unambiguously in both subunits; however, regions of local disorder are encountered at the amino termini, the carboxy

Table 1 The heavy-atom derivative preparation and refinement statistics

	No. of heavy-atom sites	Minimum resolution (Å)	R_{sym}	Resolution limits (Å)									
				40.00	23.20	11.60	7.73	5.80	4.64	3.87	3.31	2.90	
Native		2.9	0.048										
MeHg(I)	5	2.9	0.063	$\frac{f_{r.m.s.}}{E_{r.m.s.} R_c}$	2.92	3.80	3.61	3.68	2.72	2.62	2.57	3.08	
					0.34	0.36	0.37	0.35	0.41	0.53	0.53	0.46	
MeHg(II)	3	2.9	0.044	$\frac{f_{r.m.s.}}{E_{r.m.s.} R_c}$	1.25	2.58	2.65	2.92	2.86	2.16	2.18	2.09	
					0.54	0.40	0.44	0.36	0.40	0.62	0.64	0.60	
K ₂ ReCl ₆ (I)	2	3.5	0.031	$\frac{f_{r.m.s.}}{E_{r.m.s.} R_c}$	1.75	1.22	1.29	1.70	1.15	0.82	0.84		
					0.37	0.59	0.66	0.63	0.71	0.79	0.98		
K ₂ ReCl ₆ (II)	3	3.5	0.032	$\frac{f_{r.m.s.}}{E_{r.m.s.} R_c}$	1.31	1.21	1.21	1.38	1.22	1.02	0.97		
					0.76	0.57	0.69	0.64	0.72	0.90	0.83		
DMM	5	5.0	0.041	$\frac{f_{r.m.s.}}{E_{r.m.s.} R_c}$	1.33	1.78	1.82	2.28	2.22				
					0.57	0.59	0.45	0.50	0.48				
			$\langle m \rangle$	0.95	0.93	0.90	0.91	0.83	0.75	0.70	0.65	Overall	
			No. of reflections	17	150	386	634	999	1383	2011	2578	8158	0.74

Derivative soaking conditions: MeHg(I): 1 mM methylmercuri- β -mercaptoethanol, 50 mM potassium phosphate, 0.5 mM cyclic AMP, pH 8.0, 1 day. MeHg(II): 0.1 mM methylmercuri- β -mercaptoethanol, 50 mM potassium phosphate, 0.5 mM cyclic AMP, pH 8.0, 1 day. K₂ReCl₆(I): 1 mM K₂ReCl₆, 50 mM potassium phosphate, 0.5 mM cyclic AMP, 0.1 mM dithiothreitol, pH 8.0, 1–2 days. K₂ReCl₆(II): 2 mM K₂ReCl₆, 50 mM 2[N-morpholino]ethane sulphonic acid, 50 mM KCl, 0.5 mM cyclic AMP, 0.1 mM dithiothreitol, pH 6.4, 2 days. DMM: a drop of dimethylmercury was placed in capillary with mounted crystal and allowed to equilibrate for 3 days. $\langle m \rangle$ = mean figure of merit.

$$R_{\text{sym}} = \frac{\sum |I_i - \bar{I}|}{\sum \bar{I}} \quad f_{r.m.s.} = \left\{ \frac{\sum f_H^2}{n} \right\}^{1/2} \quad E_{r.m.s.} = \left\{ \frac{\sum (F_{PH} - |F_P + f_H|)^2}{n} \right\}^{1/2} \quad R_c = \frac{\sum |F_{PH} - F_P| - f_H}{\sum |F_{PH} - F_P|}$$

where I = diffraction intensity, f_H = calculated heavy-atom structure factor amplitude, F_P = structure factor amplitude of native crystals, and F_{PH} = structure factor amplitude of derivative crystals. Summations are overall reflections, n .

termini and a coiled structure in the smaller domain marked A and A'.

Approximate α -carbon positions were obtained by placing markers on map sections and measuring their coordinates. Polarity of the polypeptide chain was established from the tilt of the larger side chains on the α -helices. Although many side chains are clearly visible in this map, no attempt has yet been made to place the several fragments of amino acid sequence obtained by Schlesinger. Cyclic AMP was identified because it was not continuously connected with the polypeptide backbone, included the most dense feature in the map (presumably the phosphate) and was too large to be an amino acid side chain.

Subunit morphology

The CAP subunit is folded into two distinct structural domains (Fig. 2). The larger amino-terminal domain has approximate overall dimensions of $25 \times 30 \times 35$ Å. It contains ~135 amino acid residues and consists primarily of a pair of helices (A and B, Fig. 2) which lead into an antiparallel β -strand folded into an eight-strand β -roll; also, a 40 Å-long α -helix C connects the two domains and partially closes off one end of the interior pocket formed by the β -roll. This helix extends from the small domain to the distal extremity of the large domain rather than forming a short, proximal bridge between the two domains. The topology of the eight-strand β -roll, sometimes referred to as the 'Swiss roll', has recently been observed, with minor variations, in three protein structures: the 'shell' domain of tomato bushy stunt virus protein²², Southern bean mosaic virus protein²³ and influenza haemagglutinin⁴⁶.

The carboxy-terminal domain is smaller, with approximate dimensions $20 \times 20 \times 30$ Å. It has about 65 amino acid residues and contains three distinct α -helices, D, E and F; the D helix is ~24 Å long, the F helix ~22 Å long, whereas the E helix is

rather short. Much of the small domain consists of irregular structure, some of which is partially disordered. Although the electron density marked AA' in Fig. 1 and corresponding to the region circled in Fig. 2 is well above the background level of the solvent region, it is not sufficiently well defined to place α -carbon positions with complete assurance. Thus, the polypeptide path shown in Fig. 2 at the carboxy terminus of the D and F helices illustrates only the general folding of the subunit in these regions and the connections within the circled region may have to be altered when amino acid sequence data are fitted to the electron density map. As the number of amino acid residues in our model is very nearly that expected from amino analysis^{10,11}, no significant portion of the molecule can be missing from the model.

Although there is a distinct cleft between the two domains of the CAP subunit, noncovalent contacts occur between them that presumably stabilize the relative orientation of the domains. Both subunits contain one similar region of noncovalent contact between their large and small domains: the end loop connecting strands 4 and 5 of the β -roll in the large domain is in contact with a loop of coil in the small domain. In addition, in one of the subunits, there is a second point at which the carboxyl end of the D helix may be within noncovalent contact distance of strand 5 of the β -roll of the large domain.

Dimer morphology

As the structures of the two subunits of the CAP dimer differ, there is no unique dyad axis relating them (Figs 3–5). Qualitatively, the cleft of one subunit appears 'closed' to the point of extensive van der Waals' contact between the large and small domains while the cleft of the second subunit is 'open'; that is, the two domains have different relative orientations in the two subunits. Quantitatively, if the two large domains are super-

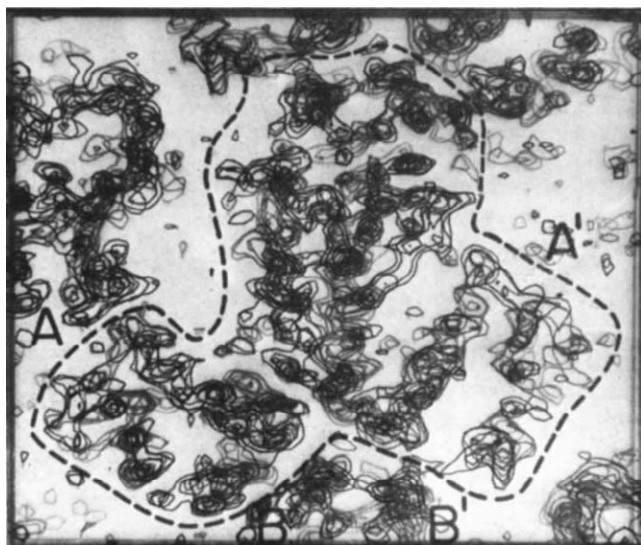


Fig. 1 Part of the electron density map, 81 Å horizontal by 71 Å vertical by 10 Å depth. Only positive contours are shown; the contour interval is approximately one-sixth the maximum peak height. The dashed line circumscribes one CAP molecule. The two oblong lobes near the bottom of the map are the small domains of the molecule; the two helices running approximately vertically near the centre of the map are the C helices. A, A', coiled structure in the smaller domain; B, B', positions of tight intermolecular contact between small domains and a large domain of a neighbouring molecule.

imposed by least-squares minimization of differences in corresponding α -carbon positions, a subsequent rotation of the small domain of one subunit by 28° is required to bring it into coincidence with the small domain of the second subunit. The 'hinge' region of the molecule appears to consist of approximately four to five amino acids in the region connecting the C helix to the small domain. This kind of conformational difference between chemically identical subunits has been widely observed, such as in yeast hexokinase²⁴, tomato bushy stunt virus coat protein²², immunoglobulins²⁵ and alcohol dehydrogenase²⁶.

Interactions between the two subunits are through direct contact along virtually the entire length of the two C helices, and contact of the β -roll of one subunit with the central region of the C helix of the other. No direct contact between the two small domains is apparent.

Although the entire dimer does not possess an overall 2-fold axis, the two pairs of domains are each related by separate rotation axes. These two rotation axes have been obtained by computing the rotation required for least-squares minimization of the difference in corresponding α -carbon positions of each domain. The results are summarized in Fig. 3. The large domains appear to be related by 2-fold symmetry (computed rotation $179.2 \pm 1.1^\circ$); the small domains may have a small distortion from dyad symmetry (computed rotation $186.7 \pm 3.7^\circ$); the angle between the rotation axis relating the large domains and that relating the small domains is 13° . The extensive contact between large domains suggests that these two domains maintain the same relative orientation in the presence and absence of various ligands. The observed asymmetry is, therefore, probably generated by an alteration in the orientation of a small domain relative to the large domains.

Cyclic AMP binding

The electron density map contains density for two cyclic AMP molecules of equal and full occupancy per CAP dimer, as anticipated from data on cyclic AMP binding to CAP in solution²⁷. The cyclic AMP is completely buried within the interior of the β -roll of the large domain. The peak of highest electron

density in the cyclic AMP molecule and the entire map is presumed to be the phosphate. Further indication of the orientation of the nucleotide is provided by crystals soaked in a solution of 8-bromoadenosine-3', 5'-cyclic monophosphate (8-Br-cyclic AMP). The cyclic AMP seems to be oriented with the phosphate bound in the loop connecting strand 6 to strand 7 in the β -roll, with the nucleotide base in proximity of the two C helices. Whether the cyclic AMP makes direct contact with one or both of the C helices, and whether the two cyclic AMP molecules per dimer are bound identically cannot be determined conclusively at this stage. The finding²⁷ that cyclic AMP shows negative cooperativity in its binding to CAP in solution may explain the formation of the asymmetric dimer observed in the crystal.

CAP domain structure and function

The crystal structure of CAP shows that the subunits consist of two distinct structural domains which may be generally true for allosterically controlled gene regulatory proteins. Limited proteolysis studies of *E. coli lac* repressor²⁸⁻³⁰, bacteriophage λ repressor³¹ and CAP^{12,13} suggest that these proteins have a small DNA-binding domain and a larger domain that binds allosteric effectors and/or holds the subunits together. In all three cases it now seems that there is little, if any, direct interaction between the two DNA-binding domains.

Proteolytic cleavage of the *lac* repressor in appropriate conditions yields two separate protein fragments^{29,30}: a tetrameric core that binds inducers, and a smaller monomeric domain having 51-59 amino acids, depending on the protease used. The small domain retains nonspecific³⁰ and operator-specific³² DNA-binding activity but does not bind inducers. Similarly, proteolytic cleavage of the λ repressor yields a dimeric core devoid of DNA-binding activity and monomeric domains which bind both nonspecific and λ operator DNA³¹.

It has been demonstrated that limited proteolysis of CAP yields an amino-terminal fragment of subunit MW 12,500, which still forms dimers and binds cyclic AMP but has lost its cyclic AMP-dependent DNA-binding activity¹². Isolation of a second intact carboxy-terminal fragment and determination of its cyclic AMP- and DNA-binding activities have not been

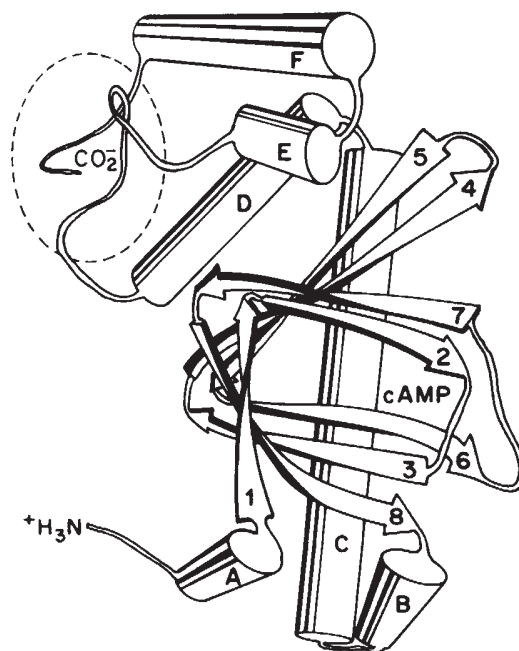


Fig. 2 Schematic drawing of the CAP promoter in the closed conformation. The numbered arrows represent β -sheet; the lettered cylinders α -helices. The binding site for cyclic AMP in the β -roll is labelled. The dashed line circumscribes the region which is not well ordered in the electron density map.

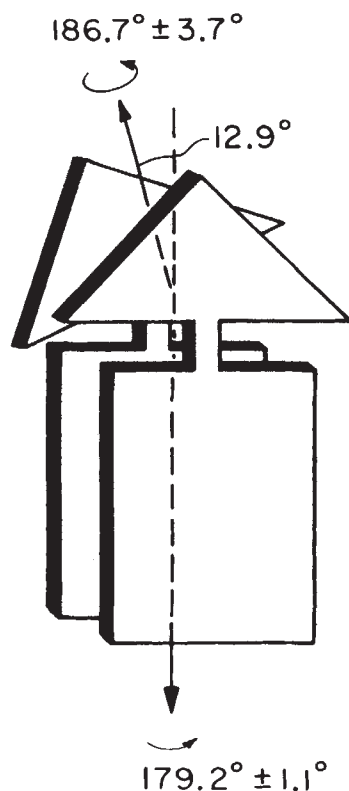


Fig. 3 Schematic diagram of the molecular symmetry of the CAP dimer. The rectangles represent the large N-terminal domains, the triangles the small C-terminal domains.

reported. Nevertheless, by analogy with the *lac* and λ repressors one might presume that the small domain of CAP binds to DNA while the larger domain binds the cyclic AMP and forms dimers.

Interaction of CAP with DNA

It has not proved possible to observe DNA binding to these crystals. Crystals soaked in solutions containing short (6–10 base pairs) oligodeoxyribonucleotides become substantially disordered. Because the small domains form tight intermolecular contacts with a large domain of a neighbouring molecule (regions B and B', Fig. 1), the most plausible DNA-binding surface of the small domains is inaccessible to oligonucleotides. Thus, direct study of DNA binding in this crystal form is probably impossible. To determine experimentally the structure of CAP bound to DNA, it will be necessary to co-crystallize CAP with an appropriate oligo-DNA.

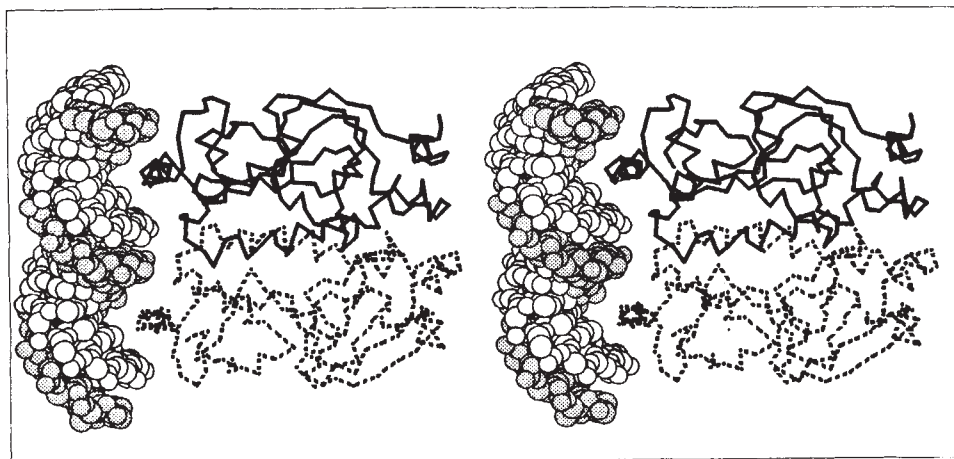
We have turned to model building to identify the DNA-binding site on CAP. Although we have not yet incorporated side chains into our CAP model, three general characteristics should be exhibited by the DNA binding site. (1) We expect a large degree of general structural complementarity between DNA and the potential binding site. (2) The correct CAP–DNA model must provide a structural basis for explaining the chemical protection^{33,34} and 'footprinting' (A. Schmitz, personal communication) data obtained on the CAP site of the *lac* operon. (3) Because the sequence of the CAP binding site at the *lac* promoter has a high degree of 2-fold symmetry, CAP probably binds to DNA with the approximate 2-fold axis that relates the two small domains coincident with the approximate 2-fold axis in the CAP site. Our conclusion is that the only model of interaction between a CAP dimer and DNA that adequately meets these three criteria requires that CAP binds to a left-handed, B-type DNA.

We have examined several possible modes of binding, and find only two ways in which a DNA molecule could interact with both small domains of a single CAP dimer with the approximate 2-fold axes of DNA and CAP coincident. Both ways involve the 22 Å-long F helices that are a dominant structural feature of the small domains, and are related by the approximate 2-fold axis between the small domains, are nearly parallel to each other and have their helix axes separated by about 34 Å. Furthermore, these helices are inclined at ~65–70° to the line connecting their centres.

The most complementary mode of interaction between the DNA and CAP is suggested by the 34 Å spacing and tilt of the 2-fold-related F α -helices. These two α -helices can fit snugly into two successive major grooves of left-handed B-DNA³⁵ (Fig. 4). In addition, the extended polypeptide chain is correctly positioned to interact with the sugar–phosphate backbone on either side of the two α -helices (Figs 4, 5c). Placing the α -helices in the major groove accounts nicely for the L8 and L29 mutations³⁶, methylation protection at the N⁷ of two pairs of guanines and the UV light-induced cross-linking of two symmetry-related thymines^{33,34}, because all these residues then interact with these helices (Fig. 5c). Furthermore, the phosphate moieties implicated by chemical modification to be interacting with CAP are almost all near the polypeptide backbone (Fig. 5c). In this way the protein can interact with at least 18–22 base pairs of DNA in a manner consistent with the footprinting data (refs 8, 34, 37 and A. Schmitz, personal communication) which show CAP protection spanning 25 base pairs.

Left-handed B-DNA has recently been shown to be a stereochemically acceptable conformation³⁵. Note that CAP cannot interact with left-handed Z-DNA^{38,39}, because the pitch is too large (45 Å) and there are no major grooves for the helices to fill. Furthermore, a purine–pyrimidine alternating sequence, requisite for formation of the Z-DNA helix, is not present in any known CAP recognition sites. Nevertheless, the known crystal

Fig. 4 Stereo drawing of the α -carbon backbone of the CAP dimer interacting with 22 base pairs of left-handed B-type DNA. One CAP subunit is drawn with dashed lines, the other subunit with solid lines. The coordinates for the DNA were derived from parameters given in ref. 37 and are presented in a space-filling representation using standard van der Waals radii. Note the extraordinary degree of general structural complementarity which extends over 18–20 base pairs (65–70 Å) of DNA.



promoter and CAP site are in negatively supercoiled rather than relaxed linear DNA (W. R. McClure, personal communication), demonstrate that CAP is most effective when it is energetically favourable to unwind, and remove negative supercoils from a closed circular DNA helix⁴¹.

Thus, although we do not know specific details of the transition, we propose that CAP traps specific regions of DNA in a left-handed helical conformation, and that the right-to-left helix transition denatures nearby regions of the helix in the promoter⁴², possibly the region specifically unwound by RNA polymerase itself⁴³.

Anderson, Ohlendorf, Takeda and Matthews⁴⁴ have recently determined the course of the polypeptide backbone of λ cro protein and find that this repressor protein has two α -helices separated by 34 Å that can fit well into the major groove of right-handed B-DNA.

Received 3 November 1980; accepted 24 February 1981.

1. Zubay, G. & Doty, P. *J. molec. Biol.* **1**, 1–20 (1959).
2. Sung, M. T. & Dixon, G. H. *Proc. natn. Acad. Sci. U.S.A.* **67**, 1616–1623 (1970).
3. Adler, K. *et al. Nature* **237**, 322–327 (1972).
4. Church, G. M., Sussman, J. L. & Kim, S. H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1458–1462 (1977).
5. Zubay, G., Schwartz, D. & Beckwith, J. *Proc. natn. Acad. Sci. U.S.A.* **66**, 104–110 (1970).
6. Epstein, W., Rothman-Denes, L. B. & Hesse, J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2300–2304 (1975).
7. deCrombrughe, B. *et al. Nature new Biol.* **231**, 139–142 (1971).
8. Ogden, S., Haggerty, D., Stoner, C. M., Kolodrubetz, D. & Schleif, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3346–3350 (1980).
9. Musso, R. E., DiLauro, R., Adhya, S. & deCrombrughe, B. *Cell* **12**, 847–854 (1977).
10. Anderson, W. B., Schneider, A. B., Emmer, M., Perlman, R. L. & Pastan, I. *J. biol. Chem.* **246**, 5929–5937 (1971).
11. Riggs, A. D., Reiness, G. & Zubay, G. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1222–1225 (1971).
12. Krakow, J. G. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2529–2533 (1973).
13. Eilen, E., Pampeno, G. & Krakow, J. S. *Biochemistry* **17**, 2469–2473 (1978).
14. Wu, C.-W. & Wu, F. Y.-H. *Biochemistry* **13**, 2573–2578 (1974).
15. Wu, F. Y.-H., Nath, K. & Wu, C.-W. *J. molec. Biol.* **100**, 459–472 (1976).
16. Eilen, E. & Krakow, J. S. *J. molec. Biol.* **114**, 47–60 (1977).
17. Kumar, S. A., Murthy, N. S. & Krakow, J. S. *FEBS Lett.* **109**, 121–124 (1980).
18. McKay, D. B. & Fried, M. G. *J. molec. Biol.* **139**, 95–96 (1980).
19. Wyckoff, H. W. *et al. J. molec. Biol.* **27**, 563–578 (1967).
20. Rossmann, M. G. *Acta crystallogr.* **A32**, 774–777 (1976).
21. Strahs, G. & Kraut, J. *J. molec. Biol.* **35**, 503–512 (1968).
22. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. *Nature* **276**, 368–373 (1978).

These data suggest the following general model for proteins that regulate gene transcription: (1) Two DNA-binding domains related by (approximate) 2-fold symmetry bind to one side of the DNA helix at the site of (approximate) 2-fold symmetry. (2) Base-specific recognition can be accomplished by α -helices of the protein binding in the major groove of B-type DNA. (3) Repressors, by binding to right-handed B-DNA, stabilize the right-handed helix and act as 'anti-melting' proteins; CAP protein, by binding to a left-handed helix, destabilizes the helix and facilitates binding and initiation by RNA polymerase.

This work was supported by USPHS grant GM-22778, NSF grant PCM-77-25111 and by postdoctoral fellowships from the American Cancer Society and US Public Health Service to D.B.M. We acknowledge assistance from Dr Harold W. Wyckoff, Paul Pepin, Gerald Johnson, John Kennedy, Art Perlo, Steve Anderson, Michael Fried and Melody Lane.

23. Abad-Zapatero, C. *et al. Nature* **286**, 33–39 (1980).
24. Bennett, W. S. Jr & Steitz, T. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4848–4852 (1978).
25. Amzel, M. & Poljak, R. J. A. *Rev. Biochem.* **48**, 961–998 (1979).
26. Eklund, H., Samama, J.-P., Wallen, L. & Branden, C.-I. *J. molec. Biol.* (in the press).
27. Takahashi, M., Blazy, B. & Baudras, A. *Biochemistry* **19**, 5124–5130 (1980).
28. Platt, T., Files, J. G. & Weber, K. *J. biol. Chem.* **248**, 110–121 (1973).
29. Geisler, N. & Weber, K. *Biochemistry* **16**, 938–943 (1977).
30. Jovin, T. M., Geisler, N. & Weber, K. *Nature* **269**, 668–672 (1977).
31. Pabo, C. O., Saver, R. T., Sturtevant, J. N. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1608–1612 (1979).
32. Ogata, R. T. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5851–5854 (1978).
33. Simpson, R. B. *Nucleic Acids Res.* **8**, 759–766 (1980).
34. Majors, J. thesis. Harvard Univ. (1979).
35. Gupta, G., Bansal, M. & Sasisekharan, V. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6486–6490 (1980).
36. Dickson, R. C., Abelson, J., Johnson, P., Reznikoff, W. S. & Barnes, W. M. *J. molec. Biol.* **111**, 65–79 (1977).
37. Taniguchi, T., O'Neill, M. & deCrombrughe, B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5090–5094 (1979).
38. Wang, A. H.-J. *et al. Nature* **282**, 680–686 (1979).
39. Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. *Nature* **286**, 567–573 (1980).
40. Sanzey, B. *J. Bact.* **138**, 40–47 (1979).
41. Cozzarelli, N. R. *Science* **207**, 953–960 (1980).
42. Dickson, R. G., Abelson, J., Barnes, W. M. & Reznikoff, N. S. *Science* **187**, 27–35 (1975).
43. Siebenlist, U. *Nature* **279**, 651–652 (1979).
44. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 754–757 (1981).
45. Arnott, S. & Hukins, D. W. L. *Biochem. biophys. Res. Commun.* **47**, 1504–1509 (1972).
46. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).

Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in *D. melanogaster*

Heli Roiha, J. Ross Miller, Lesley C. Woods & David M. Glover

Cancer Research Campaign, Eukaryotic Molecular Genetics Research Group, Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK

The two types of Drosophila melanogaster rDNA insertion occur at different sites in the 28S gene, separated by 51 nucleotide pairs. A short segment of rDNA is deleted at the site of the 5 kilobase type I insertion, whereas there is no such deletion of rDNA sequences which flank the type II insertion. Type I insertion elements in tandem arrays are flanked by very short segments of the 28S rRNA gene.

A LARGE proportion of the ribosomal genes of *Drosophila melanogaster* contain an intervening sequence in the 28S rRNA gene^{1–4}. There are two types of intervening sequence which have been termed the type I and type II insertions. The type I insertions are found in about 60% of the rDNA units of the X chromosome^{5–7} and the type II insertions in about 15% of the rDNA units of both the X and Y chromosomes^{6,8–10}. Physical mapping studies showed that these two non-homologous types of insertion are apparently located within the rDNA at identical sites, suggesting that there might be a structural feature of this region which could promote either the acquisition or loss of these elements. A high proportion of type I sequences is also

present in the chromocentral heterochromatin^{7,11}, linked together in tandem arrays¹². As far as is known, the type II insertions are only found in ribosomal gene repeats.

Intervening sequence elements have subsequently been reported in the distal part of the gene for the large rRNA of several other organisms^{13–18}. The nucleotide sequences at the junctions between the gene and the intervening sequence have been determined for the 28S gene from *Tetrahymena*¹⁹, the chloroplast rDNA from *Chlamydomonas*²⁰ and the mitochondrial rDNA of yeast²¹. In none of these cases is it possible to recognize a consensus sequence which might serve as a recognition site for enzymes which process the primary rRNA

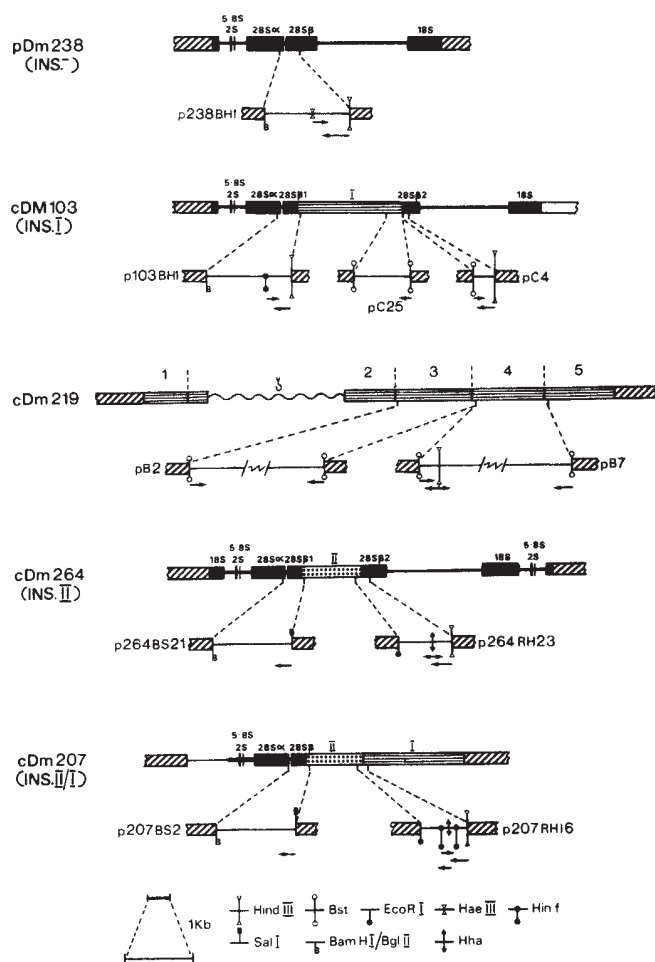


Fig. 1 Sequencing strategies. The physical maps presented show cloned fragments of chromosomal DNA which contain representative arrangements of the rDNA insertion sequences. Subcloned fragments of these DNAs which contain the junction of the insertion sequences with their flanking sequences are shown beneath each map. The construction and isolation of the original genomic clones cDm103^{31,32}, cDm219¹², cDm264¹⁰ and cDm207⁸ have been described elsewhere. The *D. melanogaster* DNA in pDm238 originates from the same set of randomly cloned segments as cDm219, cDm264 and cDm207, except that in this case a 12-kilobase *Eco*RI fragment from the original clone has been subcloned into the plasmid pBR322. All the subclones described were constructed in the appropriate sites of pBR322 as previously described¹². The exceptions are p238BH1 and p103BH1, in which the vector is pAT153³³, a non-mobilizable derivative of pBR322. The restriction site symbolized as *Bam* HI/*Bgl* II represents the product of ligating *Bam* HI and *Bgl* II termini in the vector and the *Drosophila* DNA, respectively. Only those restriction endonuclease cleavage sites used for either subcloning or sequencing are shown. The arrows beneath the maps of the subclones indicate the extent and direction of the sequences read from various restriction sites. All sequencing was performed using the techniques of Maxam and Gilbert³⁴. In each of the maps of the original clones, the ribosomal genes are indicated by filled blocks, the type I insertions by horizontal hatching and the type II insertion by dotted shading. These regions are not indicated in the subclones. The plasmid vectors are indicated by cross-hatching.

transcript, such as has been postulated for genes transcribed by RNA polymerase II²². Contrary to these cases, it is not certain whether those genes from *D. melanogaster* which contain the type I insertion sequence are transcribed into functional rRNA, as full-length transcripts corresponding to the gene and intervening sequence are extremely rare²³ and furthermore, there are rDNA units within the genome which do not contain intervening sequences.

We describe here the nucleotide sequences of the junctions between rDNA and the insertions and show that the two types of insertion occur at sites which are 51 nucleotides apart. The type I insertion occurs at an analogous site to the insertion in *Drosophila virilis* rDNA^{24,25}, but in all our clones we find that by comparison with the *D. virilis* sequence, 22 nucleotides of rDNA are deleted to the left-hand side of the insertion. There is no such deletion in the rDNA sequences flanking the type II

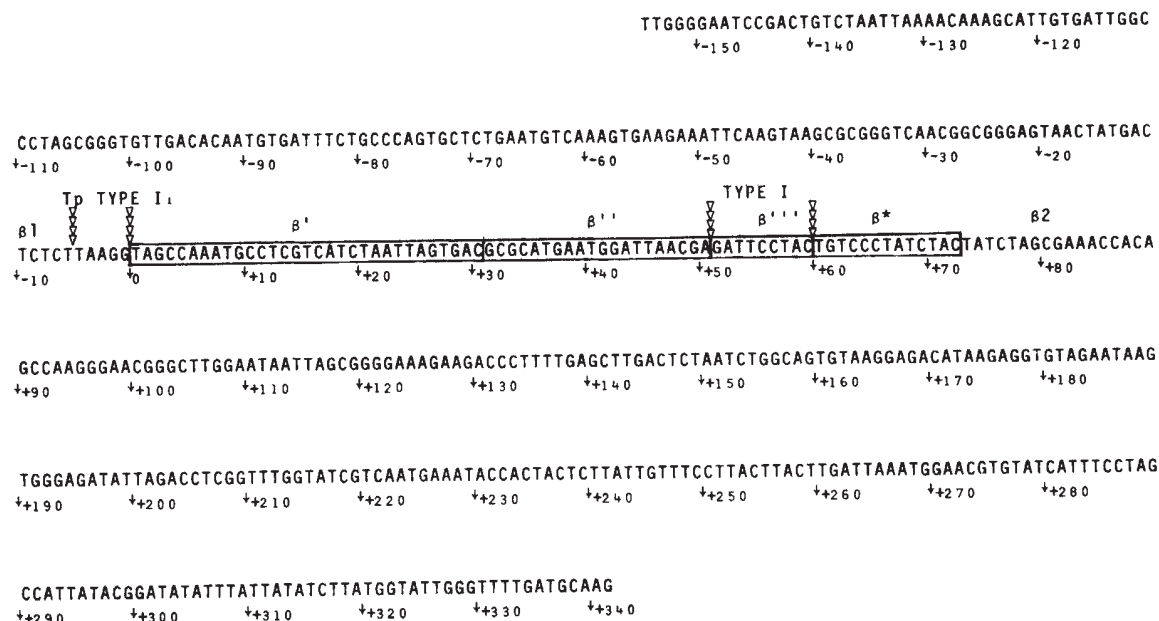


Fig. 2 Sequence of a 497-nucleotide segment of the 28S β gene. This sequence was primarily obtained from p238BH1, a subclone of an uninterrupted unit of rDNA (see Fig. 1). Plasmid DNA was labelled for sequencing either at the *Hind* III site by 'filling in' the cohesive terminus using reverse transcriptase or at the *Hae* III site using terminal transferase-mediated addition of an α -labelled ribonucleotide³⁴. The sequence of the equivalent regions in cDm103 and cDm264 have also been determined using comparable labelling techniques (see Figs 3 and 6 legends). Other than a short deletion of rDNA sequences, β'' , in cDm103 (see text) there are no differences in the segments of rDNA which have been sequenced in these plasmids. The two 'halves' of the 28S β gene on either side of the insertion are termed β 1 and β 2. We have arbitrarily begun our numbering from the position of the type II insertion which is marked with an arrow. In 28S β genes containing the type I sequence, the β'' segment is substituted by the insertion. The β^* sequence is analogous to a segment of rDNA which is duplicated in *D. virilis* rDNA units containing insertions²⁵, and which is also present in the clone cDm219 whereas β' and adjoining β 1 sequences are not. The division of the 28S β gene into these units is to facilitate comparisons between various plasmids and does not necessarily denote any functional significance of these regions. The arrow marked Tp indicates the position of the intervening sequence in a homologous region of the 28S gene of *Tetrahymena pigmentosa*¹⁹.

Differences between the junctions of the major insertion in

D. melanogaster and *D. virilis*

Rae *et al.*²⁵ have demonstrated that in *D. virilis* rDNA a 14-nucleotide sequence present only once at the site of the insertion in the uninterrupted rDNA unit is repeated on both sides of the insertion. We find that the type I insertion in cDm103 is at this same site, flanked on its right-hand side by the same 14-nucleotide sequence as found in *D. virilis* which we have called β^* (Fig. 2). The structure of the left-hand side differs from that of *D. virilis*: There is no β^* sequence and, furthermore, there is a deletion of nine residues of rDNA which we have called the β''' segment (Fig. 3).

The arrangement of the β^* sequence in *D. virilis* resembles that found at the insertion sites of transposable elements of both prokaryotic²⁶ and eukaryotic^{27,28} organisms. The occurrence of an insertion at the same site in *D. melanogaster* suggests that if the insertions arose in the rDNA of ancestral Dipteran species as a result of transposition, then the resulting unit has been reasonably stable throughout evolution. Alternatively, the duplication of the β^* sequence may be analogous to the duplication of *att* site sequences resulting from homologous recombination which occurs between the genomes of *Escherichia coli* and its lambdoid phages²⁹. In this case site-specific recombination would perhaps be a continuous process in the rDNA of *Drosophila* species. The deletion of rDNA sequences from the left-hand side of the *D. melanogaster* insertion is also analogous to the type of deletions that can be generated on excision of either prokaryotic transposable elements or lambdoid prophage. Although rDNA units containing such deletions are clearly stably maintained to some extent within the genome it remains of interest to know whether there are rDNA units within *D. melanogaster* which have the duplicated segments found in *D. virilis*. This would help answer some of the above questions.

Small rDNA segments occur in the chromocentral type I sequences

As many as 50% of the type I sequences are not inserted into rDNA but are located within the chromocentral heterochromatin and also at a single site on chromosome 4 in band 102C^{7,12}. The cloned DNA segment cDm219 originates from the chromocentral heterochromatin and contains five discrete units of the type I sequence arranged in tandem. We have determined the sequences between the junctions of the units of cDm219 following the strategy outlined in Fig. 1. The sequence at the junction of units 3 and 4 of the tandem array is given in Fig. 4, where the sequences corresponding to the left- and the right-hand ends of the rDNA insertion have positive and negative numbers, respectively. At the left hand of these units we find

20 nucleotides of rDNA, the β'' segment, the same segment of the 28S β gene which flanks the left hand of the type I insertion in rDNA (Fig. 3). At the right-hand end of these units the β^* segment flanks the insertion as in Dm103 (Fig. 3), and at the junctions between the tandemly arranged units of cDm219 the β'' segment is linked to the β^* segment. The relative positions of these small segments of the 28S β gene which flank the type I insertion of *D. melanogaster* rDNA and the insertion in *D. virilis* rDNA are shown schematically in Fig. 5. The finding of short segments of rDNA which flank units of chromocentral type I sequences strongly suggests that type I elements can be transposed from the nucleolus and that co-transposition of adjoining rDNA segments can occur during this process. The excised unit could then be amplified either by autonomous replication or by unequal exchange following integration into its new chromosomal site. If the insertion elements of both *D. virilis* and *D. melanogaster* are indeed mobile, these observations would suggest that only the right-hand β^* element is required for transposition. Perhaps it is significant in this respect that within the rDNA sequence to the right of β^* in both *D. virilis* and *D. melanogaster* there is a direct duplication of part of the β^* sequence (residues 66–79, Fig. 2).

The continuous sequence given in Fig. 4 is that of the junctions of the cDm219 subclones pB2 and pB7. We have also sequenced about 200 residues on either side of the junctions in other subcloned type I units of cDm219, and have found no differences between these junctions and the sequence shown in Fig. 4. Figure 4 also compares the sequences from the left and right ends of the rDNA insertion in cDm103. Within this limited sequence there seems to be a significant difference in the degree of conservation between the left and right ends of these insertion elements. The only base difference within a comparable 226-nucleotide sequence from the right-hand end is the presence of an extra dA residue in cDm219. At the left end, however, there are 19 nucleotides which differ within a comparable 231-nucleotide sequence. There is similar variability in an equivalent sequence from the left-hand end of the type I insertion found in cDm207 (see below); this sequence has 10 single-base differences when compared with this Dm219 element and 20 compared with cDm103. There is also a 22-nucleotide deletion in this region of cDm207. In this respect, we note that although there is considerable length variation of the type I sequences, it is the sequences from the right-hand side of the unit which are preserved^{3,6}. It will be interesting to examine the nature of junctions of these shorter type I elements with adjoining rDNA.

Junctions of the type II insertions

About 15% of the *D. melanogaster* rDNA units have insertions with cleavage sites for *Eco*RI^{6,8}. We have characterized several cloned segments of rDNA containing complete type II sequences and find no homology with the type I sequences¹⁰. We chose to determine the sequence at the junctions of the type II insertion of cDm264 because this insertion has the most commonly observed length and restriction endonuclease cleavage pattern. We find that there is a direct sequence repeat at the left-hand junction of the type II insertion with the 28S β gene. The repeated sequences, GGGGAG, are contiguous and are formed partly from rDNA sequence and partly from insertion sequence (Fig. 6a). This is quite distinct from the duplication of the β^* sequence flanking the *D. virilis* insertion. The right-hand junction of the type II insertion shows a striking tract of 22 dA residues, which are followed by the β' , β'' , β''' , β^* and β_2 segments of the 28S gene (Fig. 6b, c). Unlike the case of rDNA containing the type I insertion, there is no deletion of adjacent rDNA sequences.

The deletion in Dm103 makes it extremely unlikely that units of this type are ever used in the production of mature rRNA. This is in accord with the findings that large nuclear transcripts containing type I sequences are extremely rare, and that the most abundant transcript is a cytoplasmic molecule of 1 kilobase, present in about 50 copies per cell²³. Chooi³⁰, however, has described the presence of long transcription units in actively

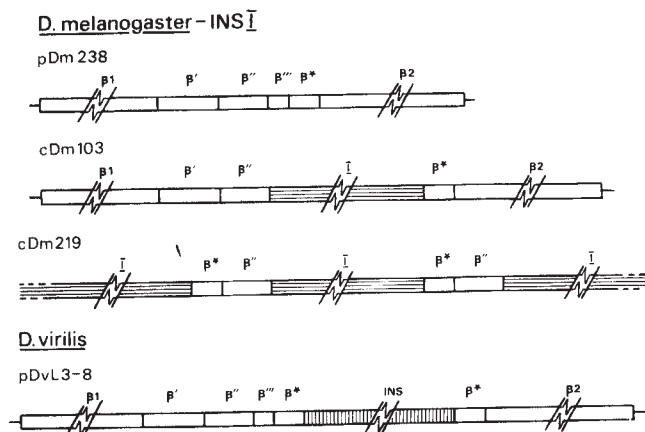


Fig. 5 A schematic comparison of the segments of the 28S β gene which flank the type I insertions in cDm103 and cDm219 with those flanking the insertion in *D. virilis* rDNA.

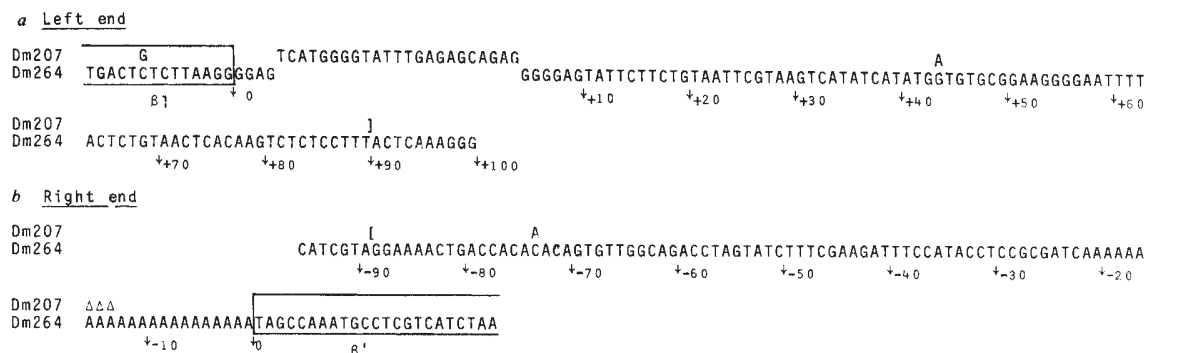


Fig. 6 Sequences at the junctions of the type II insertions within rDNA. *a*, The sequence of the left end of the type II sequences in cDm264 and cDm207 comes from the subclones p264BS21 and p207BS2, respectively. The restriction fragments were labelled at the *Sal*I site by 'filling-in' the *Sal*I cohesive end using reverse transcriptase. Only the regions in which the cDm207 sequence differs from the cDm264 sequence are shown. *b*, The sequence of the right end of the type II sequence in cDm264 and cDm207 from the subclones p264RH23 and p207RH16, respectively. *Hind*III and *Hinf* termini were 3'-labelled using reverse transcriptase and the *Hae*III termini 3'-labelled using terminal transferase. The single bases of the cDm207 sequence which differ from cDm264 are indicated. *c*, Schematic diagram of how these rDNA segments differ to the right of the type II insertion. The segments β' and β'' , defined in Fig. 2, are found in each plasmid. In cDm264, β' is contiguous with the remainder of the 28S β gene, whereas in cDm207 it is linked to the left end of the type I sequence. This type I sequence in cDm207 differs from those of cDm219 and cDm103 at the sites indicated in Fig. 4.

transcribed chromatin which could represent transcription of rDNA units containing insertions. The distribution of these long transcription units correlates with the distribution of type II sequences in the rDNA of the X and Y chromosomes. It is possible, therefore, that rDNA containing type II insertions has a similar transcriptional pattern to that found in *Tetrahymena* rDNA where the insertion sequence is present in the primary transcript and is subsequently removed. Moreover an A + T-rich region seems to be a common feature of the right-hand junction of the insertions in the rDNA of *Tetrahymena*¹⁴, *Chlamydomonas* chloroplasts²⁰ and yeast mitochondria²¹, for which the A + T content of the rightmost 25 nucleotide pairs is 72, 80 and 76%, respectively. Indeed, it has been postulated that A + T-rich regions might be an important feature of other processing sites in ribosomal transcripts.

A 28S β rRNA gene containing both insertions

We have previously found that the clone cDm207 contained both types of insertion and that within our mapping resolution the junction of type II with type I sequences corresponded to the right and left ends of these respective sequences, as they are found singly within rDNA¹⁰. Nucleotide sequencing indicates that the type II element occurs at exactly the same site as in cDm264 (Fig. 6*a*, *b*). Interestingly, the type II sequence in cDm207 differs from that of cDm264 by an additional 23 nucleotides which occur between the two repeated hexameric sequences. A tract of 19 dA residues occurs at the right-hand end of the type II insertion in cDm207. However, it is linked to only 51 nucleotides of rDNA, corresponding to the β' and β'' segments, which intervene between the sites of the two insertions. The junction between the β'' segment and the type I insertion is identical to that found in cDm103 and in the units of cDm219. The comparative structures of cDm207 and cDm264 are indicated schematically in Fig. 6*d*.

We previously noted that the restriction map of the type I insertion in cDm207 is more similar to that of a Dm219 unit than of the Dm103 rDNA insertion. A sequence comparison within a limited segment of these DNAs is consistent with this observation (see above and Fig. 4). The chromosomal segment Dm207 also contains an unusual interruption of the internal transcribed spacer region (see Fig. 1), and this has previously led us to propose that it was the product of a recombinational event between an rDNA unit containing type II sequences and type I sequences elsewhere in the genome. That recombinational events occur to produce rDNA segments containing both types

of insertion is to be expected because the insertions occur at different sites. Several such clones have been isolated by ourselves and others and we expect that sequencing will show such clones to have analogous organizational features to those which occur between the two types of insertion in Dm207.

This work was supported by project grants from the MRC. D.M.G. holds a career development award from the Cancer Research Campaign. We thank Simon Kidd for his interest in this work, for providing us with data on the positioning of restriction endonuclease cleavage sites and for the construction of several of the subcloned plasmids described here. We also thank Therese Anderton for help with sequencing techniques and Sue Hayman for typing the manuscript, especially the figures.

Note added in proof: We have recently sequenced the junctions between rDNA and a 1.0 kilobase type I insertion. This insertion is flanked by duplicated β^* segments and there is no deletion of rDNA sequences.

Received 28 November 1980; accepted 5 February 1981.

- Glover, D. M. & Hogness, D. S. *Cell* **10**, 167-176 (1977).
- White, R. L. & Hogness, D. S. *Cell* **10**, 177-192 (1977).
- Wellauer, P. K. & Dawid, I. B. *Cell* **10**, 193-212 (1977).
- Pellegrini, M., Manning, J. & Davidson, N. *Cell* **10**, 213-224 (1977).
- Tartof, K. D. & Dawid, I. B. *Nature* **263**, 27-30 (1976).
- Wellauer, P. K., Dawid, I. B. & Tartof, K. D. *Cell* **14**, 269-278 (1978).
- Peacock, W. J., Appels, R., Endow, S. & Glover, D. M. *Genet. Res.* **38**, (1980).
- Glover, D. M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4932-4936 (1977).
- Dawid, I. B., Wellauer, P. K. & Long E. O. *J. molec. Biol.* **126**, 749-768 (1978).
- Roiha, H. & Glover, D. M. *J. molec. Biol.* **140**, 341-355 (1980).
- Dawid, I. B. & Botchan, P. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4233-4237 (1977).
- Kidd, S. J. & Glover, D. M. *Cell* **19**, 103-119 (1980).
- Wild, M. A. & Gall, J. G. *Cell* **16**, 565-573 (1979).
- Din, N. & Engberg, J. *J. molec. Biol.* **134**, 555-574 (1979).
- Gubler, U., Wyler, T. & Braun, R. *FEBS Lett.* **100**, 347-350 (1979).
- Campbell, G. R., Littau, V. C., Melera, P. N., Allfrey, V. G. & Johnson, E. M. *Nucleic Acids Res.* **6**, 1433-1447 (1979).
- Devries, H., Dejonge, J. C., Bakker, M., Meurs, H. & Kroon, A. *Nucleic Acids Res.* **6**, 1791-1803 (1979).
- Hahn, U., Lazarus, C. M., Lunsdorf, H. & Kuntzel, H. *Cell* **17**, 583-595 (1979).
- Wild, M. A. & Sommer, M. *Nature* **283**, 693-694 (1980).
- Allet, B. & Rachaix, J. *-D. Cell* **18**, 55-60 (1979).
- Dujon, B. *Cell* **20**, 185-197 (1980).
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853-4857 (1978).
- Long, E. O. & Dawid, I. B. *Cell* **18**, 1185-1196 (1979).
- Barnett, T. & Rae, P. M. *Cell* **16**, 763-775 (1979).
- Rae, P. M., Kohorn, B. D. & Wade, R. P. *Nucleic Acids Res.* **8**, 3491-3504 (1980).
- Calos, M. P. & Miller, J. H. *Cell* **20**, 579-595 (1980).
- Dunsmuir, P., Brorein, W. J., Simon, M. A. & Rubin, G. M. *Cell* **21**, 575-580 (1980).
- Farabaugh, P. J. & Fink, G. R. *Nature* **286**, 352-356 (1980).
- Landy, A. & Ross, W. *Science* **197**, 1147-1160 (1977).
- Choi, W. Y. *Chromosoma* **74**, 57-74 (1979).
- Glover, D. M., White, R. L., Finnegan, D. J. & Hogness, D. S. *Cell* **5**, 149-157 (1975).
- Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961-3965 (1975).
- Twigg, A. J. & Sherratt, D. *Nature* **283**, 216-218 (1980).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

Structure of the cro repressor from bacteriophage λ and its interaction with DNA

W. F. Anderson^{*†}, D. H. Ohlendorf^{*}, Y. Takeda[‡] & B. W. Matthews^{*}

^{*} Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

[†] MRC Group on Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

[‡] Chemistry Department, University of Maryland, Baltimore County, Catonsville, Maryland 21228, USA

The three-dimensional structure of the 66-amino acid cro repressor protein of bacteriophage λ suggests how it binds to its operator DNA. We propose that a dimer of cro protein is bound to the B-form of DNA with the 2-fold axis of the dimer coincident with the 2-fold axis of DNA. A pair of 2-fold-related α -helices of the repressor, lying within successive major grooves of the DNA, seem to be a major determinant in recognition and binding. In addition, the C-terminal residues of the protein, some of which are disordered in the absence of DNA, appear to contribute to the binding.

INTERACTIONS between proteins and nucleic acids are of central importance in molecular biology. Some proteins bind single-stranded polynucleotides while others recognize double-stranded ones. None of these interactions is well understood in molecular terms. In particular, the way in which proteins recognize specific base sequences within double-stranded DNA has remained enigmatic.

The cro protein from bacteriophage λ (hereafter called 'cro') is a repressor molecule which recognizes and binds tightly certain sequence-specific operator regions within the DNA of the bacteriophage. Cro is the smallest repressor protein that has been characterized, and is of interest both because of its role in the delicately balanced regulatory system of bacteriophage λ , and as a general model for protein-DNA interaction.

We describe here the three-dimensional structure of the cro repressor and its apparent mode of interaction with operator DNA. The postulated interaction between repressor and operator is consistent with a variety of chemical and genetic evidence, and may provide a general model for the interaction of proteins with helical DNA.

The bacteriophage λ genome encodes two repressor proteins that play essential, contrasting parts in regulating the development of the phage. The product of the *cI* gene, the *cI* or λ repressor ('*cI*'), is required for the maintenance of the lysogenic state, whereas the product of the *cro* gene is necessary during the lytic cycle of phage development (for reviews see refs 1-4, 33, 34). Both the *cro* and the *cI* repressors compete for and bind two operator regions within the phage genome. These operator regions each consist of three binding sites for the repressor molecules, and the respective repressors have different relative affinities for each of the sites within the operator regions^{5,6}. The use of these binding sites by *cro* and *cI* provides for a switch between the lysogenic and lytic pathway. The *cro/cI* interaction is the clearest example of such a developmental switch³³ and the structural results presented here begin an understanding of this switch at the molecular level.

The *cro* protein is well characterized^{7,8}. It is a small basic protein 66-amino acid residues long, with a monomer molecular weight of 7,351. Both the amino acid sequence of the protein⁹ and the nucleotide sequence of the *cro* gene have been determined¹⁰.

Structure determination

The structure was determined by the method of isomorphous replacement¹¹⁻¹³. Protein purification and crystallization from 1.1 to 1.3 M phosphate were as described previously^{8,14}. The space group is R32 and the cell dimensions in the hexagonal system $a = b = 91.9$ Å, $c = 268.9$ Å.

Potential heavy-atom derivatives were initially screened by precession photography, and, later, by using a diffractometer to collect 5-Å-resolution data sets. The first heavy-atom derivative to be interpreted was obtained using PtCl_4^{2-} and was initially solved by a 'brute force' approach in which all possible indi-

vidual sites, and then combinations of sites, were explored (ref. 15 and unpublished results of B.W.M.). Using the phases obtained from this platinum derivative, other potential derivatives were then evaluated in the normal way^{12,13}. In the end, five derivatives were selected for high-resolution data collection (Table 1). All the derivatives have multiple binding sites, as expected with multiple copies of the protein present in the crystallographic asymmetric unit (see below). Higher-resolution data were collected by oscillation photography¹⁶ which permitted collection of data sets to a nominal resolution of 2.8 Å or better from a single crystal. The crystals often grow as long triangular prisms¹⁴, in which case they can be translated between series of exposures. Some data collection statistics are included in Table 1.

Two preliminary electron density maps to 5 Å resolution were calculated, one using the data collected by diffractometry, and the other using a subset of the higher-resolution oscillation data. The two maps were quite similar, and, as discussed below, were important in working out the arrangement of the *cro* monomers within the crystallographic asymmetric unit.

Molecular symmetry

The initial analysis of the *cro* protein crystals showed them to contain multiple copies of the protein in the asymmetric unit. On the basis of the density measurement of two cross-linked crystals, there seemed to be six monomers, that is, three dimers, per asymmetric unit¹⁴. However, examination of the 5 Å-resolution electron density map revealed that there were four monomers per asymmetric unit, not six. The error in the previous estimate could be due either to the presence of the cross-linking agent or to errors in density due to the limited amount of material available.

We attempted to determine the relation between the monomers in two ways—by inspection of the 5-Å-resolution electron density maps, and by consideration of the relation between the sites of heavy-atom binding. Initial inspection of the 5-Å-resolution map revealed an obvious 2-fold symmetry

Table 1 Heavy-atom derivatives

Derivative	No. of sites	R_{Merge} (%)	R_{Der} (%)	Reflections above 3σ	R_{wc} (%)	$\langle f_{\text{H}} \rangle$	E
Native	—	9.5	—	9184	—	—	—
PtCl_4^{2-}	6	9.6	14.5	5068	30.5	64.3	54.5
Hg^{2+}	16	8.9	12.9	3616	44.9	113.4	72.4
Au^{2+}	16	8.7	14.0	7273	47.1	59.5	46.8
EMTS	12	0.3	11.0	7856	48.7	39.0	32.2
Mersalyl	13	10.2	14.0	5985	46.1	62.5	51.5

Overall figure of merit¹⁸ to 2.8 Å resolution, 0.45. $R_{\text{Merge}} = 100 \sum (I_i - \langle I \rangle) / \sum I_i$, $R_{\text{Der}} = 200 \sum |I_{\text{PH}} - I_{\text{P}}| / \sum (I_{\text{PH}} + I_{\text{P}})$, $R_{\text{wc}} = 100 \sum \omega |F_{\text{PH}} - F_{\text{P}}| - f_{\text{H}}| / \sum \omega |F_{\text{PH}} - F_{\text{P}}|$. $\langle f_{\text{H}} \rangle$ is the mean heavy-atom scattering, E the average lack of closure for centric reflections¹⁸. EMTS, ethyl mercury thiosalicylic acid. Hg^{2+} data set collected by precession photography.

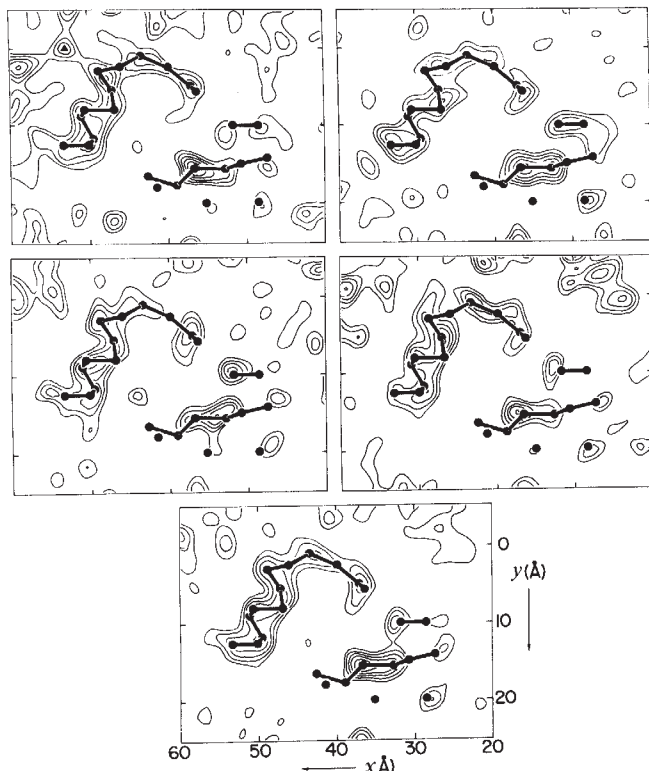


Fig. 1 Corresponding electron density sections through each of the four cro molecules in the crystallographic asymmetric unit, transformed into a common coordinate frame. Clockwise from top left the density is for molecules O, A, B and C, with the averaged density below. α -carbon atoms within ± 2.5 Å of the section are indicated by the solid circles.

axis lying approximately in the plane $z = 1/12$. This 2-fold axis also related, in pairs, most of the major heavy-atom sites. Further analysis of the 5-Å map, facilitated by the use of an MMS-X computer graphics system¹⁷, suggested that there were two additional 2-fold axes inclined to the first. The positions of these local symmetry elements were confirmed by calculating correlation coefficients between the electron density of the individual monomers, rotated and translated to superimpose one monomer on another (Table 2). The four monomers are related in pairs by mutually perpendicular intersecting (or nearly intersecting) 2-fold axes. In other words, the four monomers are arranged in the crystal with local 222-point symmetry. The symmetry of the molecules in the crystal does not itself imply anything about the state of aggregation in solution. In principle, the protein could exist as individual monomers, as dimers with 2-fold symmetry, or as tetramers. However, the specific interactions which we observed between the four monomers (see below) suggests that cro may exist in solution as a tetramer. The extent to which the molecular arrangement may depart from exact 222 symmetry, and the potential biological relevance of such deviations from exact symmetry, remain to be determined.

Most, but not all, of the heavy-atom sites are related to each other by the 222 local symmetry elements. For each 2-fold axis, related pairs of sites can be superimposed within ~ 1.9 Å (Table 2). Although the 2-fold symmetry was used as an additional check in the interpretation of the heavy-atom derivatives, it was not imposed on the heavy-atom coordinates.

The electron density map

An electron density map was calculated to a nominal resolution of 2.8 Å using the five heavy-atom derivatives described in Table 1 (ref. 18). Although the phase angles are well determined to about 3.5 Å resolution, their accuracy at higher angles is limited by the scattering intensity of the derivative crystals, which becomes very weak at higher resolution, and also, in part, by non-isomorphism.

To improve the accuracy of the electron density map and facilitate its initial interpretation, the density of the four monomers in the asymmetric unit was averaged (Table 2, Fig. 1). If the assumed transformations are correct and all four monomers are, in fact, identical, the averaging process should improve the accuracy of the electron density map. In practice, the averaging process clearly improved the overall 'cleanliness' of the map, but also resulted in a loss of detail and reduction in the effective resolution (Fig. 1). Presumably, application of more sophisticated averaging techniques¹⁹ and refinement of the individual monomers will lead to a more precise description of the structure.

The averaged electron density map was displayed both as a 'mini-map' and on a scale of 2 cm = 1 Å in an optical comparator^{20,21}. Several helices and a region of β -sheet were immediately apparent, and the course of the polypeptide backbone could be readily followed from one end to the other. The direction of the polypeptide chain was initially deduced from the observation that the major platinum binding site was at a position 12 residues from one terminus, obviously corresponding to Met 12. From this point the known amino acid sequence of the protein^{9,10} could be compared with the electron density map. Although the interpretation of the electron density map is fully consistent with the sequence, note that the limited resolution of the current map, together with the averaging process described above, results in poorly defined density for the respective side chains. The backbone density is strong and unambiguous from the amino terminus to residue Lys 62, but at this point the density becomes very weak, indicating that the four carboxy-terminal residues are disordered in the crystals. This appears to be the case for each of the four cro monomers.

Conformation of the cro repressor

The conformation of a cro monomer is illustrated in Fig. 2. The structure is very simple, consisting of three strands of antiparallel β -sheet (residues 2–6, 39–45 and 48–55), and three α -helices (residues 7–14, 15–23 and 27–36). As described below, the C-terminal residues form an extended 'arm' (Fig. 2) which interacts with another monomer and seems to participate in DNA binding.

The detailed arrangement of the four repressor monomers in the crystals is shown in Fig. 7. The respective local 2-fold symmetry axes relating the monomers are designated P, Q and R, and the monomers themselves O ('original'), A, B and C (Table 2). It has been assumed for some time that the cro

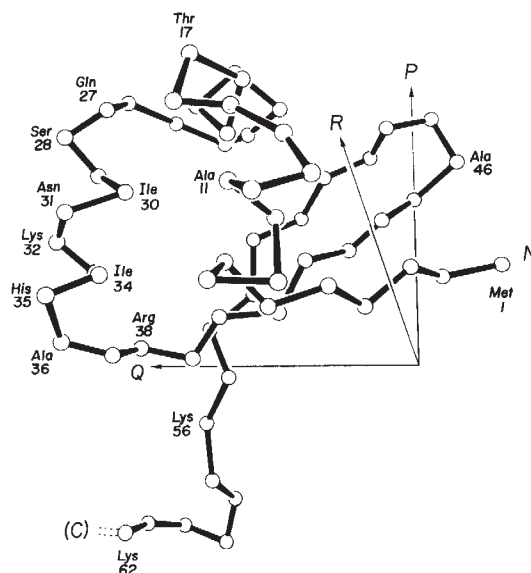


Fig. 2 Backbone of monomer O of the cro repressor protein. As discussed in the text, the four C-terminal residues 63–66 are disordered. The local orthogonal 2-fold symmetry axes P, Q and R relate the molecule shown to monomers A, B and C, respectively.

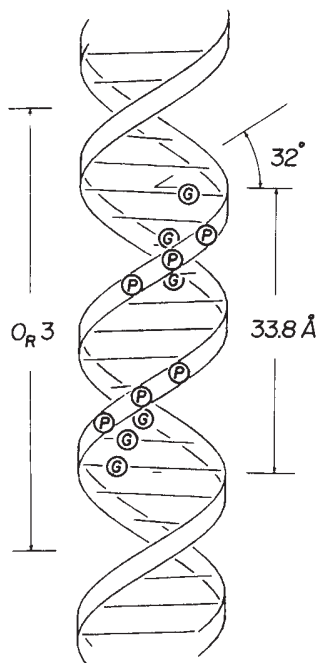


Fig. 3 DNA operator O_{R3} , drawn in the B form (after Ptashne *et al.*⁴). Ethylation of the phosphates, labelled P, hinders cro repressor binding; binding of cro protects the N^7 positions of guanines, labelled G, from methylation by dimethyl sulphate. The figure is based on experiments of Johnson, Ptashne and collaborators.^{4,5,35}

repressor exists as a dimer in physiological conditions⁸. Therefore, we expected the four cro monomers in the crystal to consist of two dimers. However, inspection of the molecular packing does not obviously support this assumption. The monomers O and B (and, similarly A and C) are related by the Q axis, and have extensive interactions between their C-terminal arms, each of which lies against, and is, in part, hydrogen bonded to, its 2-fold-related partner. It seems likely that this interaction persists in solution. However, there are also specific interactions about the R symmetry axis, where the polypeptide backbone of one monomer hydrogen bonds to the 2-fold-related backbone of the other monomer so as to expand the three-strand β -sheet of the monomer into a six-strand antiparallel sheet in the dimer. In addition, the amino terminus of monomer O is linked by an ionic interaction to Glu 53 of monomer A, with similar interactions for the other amino termini. The four amino-terminal methionines are grouped together and participate in a distinct hydrophobic region at the centre of the cro tetramer. Thus, although the packing arrangement seen in the crystal does not prove that cro exists as a tetramer in solution, it does suggest this. Whether cro exists as a dimer or tetramer has no bearing on the proposed model for DNA binding.

Interaction of cro with DNA

Both cI and cro repressors compete for three sites within the left and the right operators of bacteriophage λ . In the right operator there are three binding sites (O_{R1} , O_{R2} and O_{R3}) each consisting of 17-base pair regions separated by spacers of 6 or 7 base pairs²². The sequences of the three 17-base pair regions are similar to each other, but not identical, and, in addition, the sequence within each region is approximately symmetric. Mutations within the 17-base pair regions reduce the binding affinity of both cro and cI, whereas mutations in the spacer regions have no effect on binding (refs 4, 6, 23, 24 and refs therein). Chemical probe experiments^{4,5,35} indicate that both cro and cI repressors bind primarily along one face of the DNA double helix and protect many of the same groups. The results of such experiments for cro binding to O_{R3} are illustrated in Fig. 3. Methylation by dimethyl sulphate of the N^7 of six guanines in the major groove is prevented by the presence of cro, but

methylation of N^3 of adenine, exposed in the minor groove of the DNA, is not prevented. Furthermore, ethylation of any of the six phosphates shown in Fig. 3 interferes with cro binding⁴. The pattern for cI is very similar, but is slightly more extended, and includes four additional phosphates. Such experiments suggest that both cro and cI contact the DNA in the major, but not the minor, groove^{4,5}. Furthermore, it seems likely that both repressors bind to the DNA with two symmetry-related subunits of the protein in contact with the two symmetry-related halves of the operator, and the 2-fold axis of the protein coincident with the 2-fold axis of the DNA. In this way the complex of bound repressor and operator would have a common 2-fold symmetry axis. There would, of course, be slight deviations from exact 2-fold symmetry because of the inexact symmetry of the DNA base sequence, and also, possibly, due to the protein.

To determine possible modes of binding of cro to DNA, we looked for a feature of the cro monomer which, together with its 2-fold related mate, would form a structure complementary to that of the operator DNA, and which, when aligned on the DNA, would be consistent with the protection experiments summarized in Fig. 3. An obvious candidate is the helix from residues Gln 27 to Ala 36. This helix lies along the surface of the protein, with isoleucines 30 and 34 contributing to the hydrophobic interior of the monomer, but with the helix otherwise protruding from the surface of the protein (Fig. 2). The corresponding helix in monomer B, related by rotation about axis Q (Fig. 4) is parallel, and the respective helices have a centre-to-centre distance of 34 Å, the same as the separation between successive major grooves of B-form DNA (Fig. 3). Furthermore, the angle made between the axes of the two α -helices and the line connecting their midpoints is equal to the angle between the major groove of the DNA and its long axis (Figs 3, 4). The sense of the respective angles is such that the two helices can be accommodated, in a very natural way, within successive major grooves of DNA in its normal Watson-Crick B conformation^{25,26}.

This hypothetical arrangement was checked by fitting a model of the cro molecule to that of DNA. We required that the 2-fold axis relating the cro monomers had to coincide with the 2-fold axis through the centre of the DNA operator. Thus, the fitting of cro to the DNA is very restrictive, allowing only two degrees of freedom—one a rotation of cro relative to the DNA, about the common symmetry axis; the second a translation of cro along the symmetry axis, which varies the distance between cro and the long axis of the DNA. The results of this first, albeit crude, fitting process are shown in Figs 5, 6 and 7.

The complementarity between protein and DNA is striking. In addition to the overall correspondence between the respective backbones of the DNA and protein, residues Ser 28, Ala 29, Asn 31, Lys 32, Ala 33 and His 35 of the above-mentioned α -helix are in a position to interact either with the exposed hydrogen bonding groups of the major groove, conferring specificity, or with the negatively charged phosphates, promoting generalized binding. Nearby residues, including Arg 38 and Lys 39, are also suitably located to contribute to binding and recognition. Close to the 2-fold symmetry axis, the sequences Glu 54–Val 55–Lys 56 of the two monomers form a pair of antiparallel β -sheet strands parallel to the minor groove which possibly interact with the DNA in the manner proposed

Table 2 Relations between the four cro monomers

Symmetry axis	Pairs of heavy-atom sites	r.m.s. error in heavy-atom sites	Monomers related	Electron density correlation
P	26	1.80	O,A	0.58
Q	26	1.97	O,B	0.54
R	25	1.94	O,C	0.55

$C_{ij} = \sum \rho_i \rho_j / (\sum \rho_i^2 \cdot \sum \rho_j^2)^{1/2}$ where the summations are over all values, ρ_i , of the 2.8 Å-resolution electron density map within 15 Å of the centre of the respective monomers.

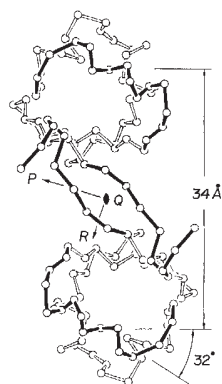


Fig. 4 Dimer of two cro molecules viewed along the Q symmetry axis. The P and R symmetry axes are also shown. Regions of the protein backbone closest to the viewer, and presumed to interact with DNA, are drawn solid. These regions include two α -helices, 34 Å apart, inclined at 32°, and also a pair of extended strands close to the Q-axis.

by Church *et al.*²⁷. The four C-terminal residues of the 2-fold-related cro monomers, which are disordered in the crystals, and also, presumably, in solution, could also interact with the DNA (Figs 5, 6, 7). Such interactions could include Lys 62 and Lys 63. Presumably these C-terminal arms, which extend in solution like 'feelers', occupy well defined positions when complexed with the DNA. The finer details of the interactions between the cro repressor and its specific operator sequences are now being explored by more careful model building, and will be described elsewhere.

The proposed binding of cro to DNA explains many of the known properties of the complex. A tetramer or a dimer of cro bound to DNA extends over exactly 17 base pairs, the length of one operator (Figs 3, 5). The mode of interaction between protein and DNA is consistent with the approximate 2-fold symmetry of the operator sequence, and the symmetry of the protein. Also, the proposed binding explains, within the present level of detail, the chemical protection experiments (Figs 3, 6). Furthermore, there is a space of ~20 Å between cro molecules bound to adjacent operators, consistent with non-cooperativity of binding²⁸.

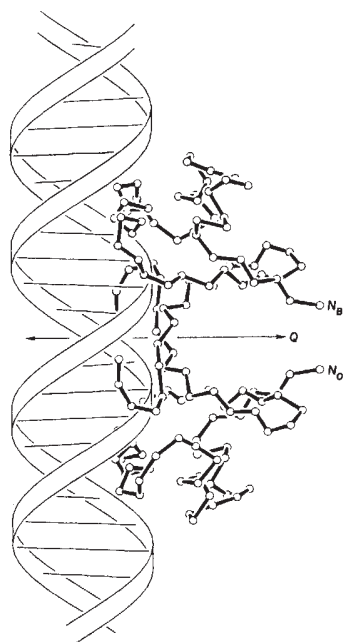


Fig. 5 Presumed interaction of cro repressor with DNA. Two monomers of cro (O and B), related by 2-fold symmetry axis Q, interact with the DNA so that axis Q coincides with the 2-fold symmetry axis of the DNA. The DNA has been rotated 90° relative to that in Fig. 3 so that its 2-fold symmetry axis is in the plane of the paper. The respective amino termini of the two cro molecules are labelled N_O and N_B . A pair of 2-fold-related α -helices occupy successive major grooves of the DNA and, closer to the symmetry axis, two extended polypeptide strands run parallel to the backbone of the DNA.

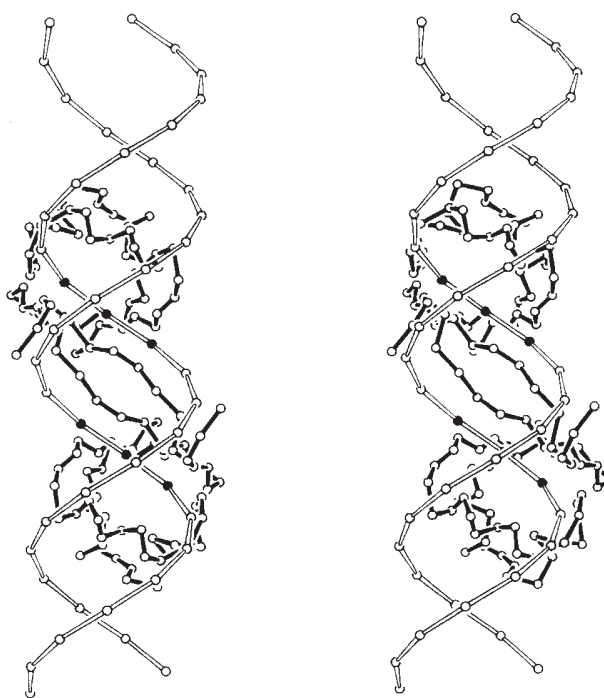


Fig. 6 Stereo view of the presumed interaction between two molecules of cro repressor and DNA, viewed along their common 2-fold symmetry axis. The open circles along the DNA backbone indicate the positions of the phosphates; those phosphates whose ethylation hinders cro binding (Fig. 3) are drawn solid.

Conclusions

This is the first structure determination of a specific DNA repressor, and the results obtained here may be of general relevance for interactions between proteins and DNA.

A principal feature of the proposed binding of cro to DNA is the use of symmetry. The 2-fold symmetry axis of the repressor coincides with the 2-fold symmetry axis of the DNA operator. This type of interaction is expected to be of general significance for proteins which recognize symmetric, sequence-specific regions of the DNA. Departures from exact symmetry could be important in 'fine tuning' of the strength of the DNA-protein complex. Our results indicate that the symmetric sequences often used for specific DNA-protein interactions are a consequence of the oligomeric structure of the regulatory proteins rather than the source of a special type of DNA structure.

Also, the α -helical conformation of proteins may presumably be used in both specific and nonspecific DNA binding. Sung and Dixon²⁹ were the first to propose that the histone H4 N-terminus adopts a helical conformation when interacting with chromatin nucleic acid. A related model has been proposed for *lac* repressor³⁰. Also, Warrant and Kim³¹ have presented direct evidence showing that protamine changes to an α -helical conformation on binding to a double-helical portion of crystalline tRNA. The cro repressor structure supports the notion that the binding of α -helices in the major groove of DNA may be a principal determinant in nonspecific protein binding to DNA, and is also of importance in specific recognition.

The structure of cro, and its complementarity to the B-form of DNA, suggest that the DNA retains the right-handed B-conformation when the repressor is bound. Of course, slight deformations of the DNA or protein could well occur on binding, and may be suggested by detailed model-building experiments, but the present results indicate that large adjustments in the DNA structure are not required to accommodate the protein. In contrast, McKay and Steitz³² have recently proposed that the cyclic AMP receptor protein of *Escherichia coli*, which promotes the action of RNA polymerase, binds to left-handed DNA. In the case of the cro repressor, the protein is located

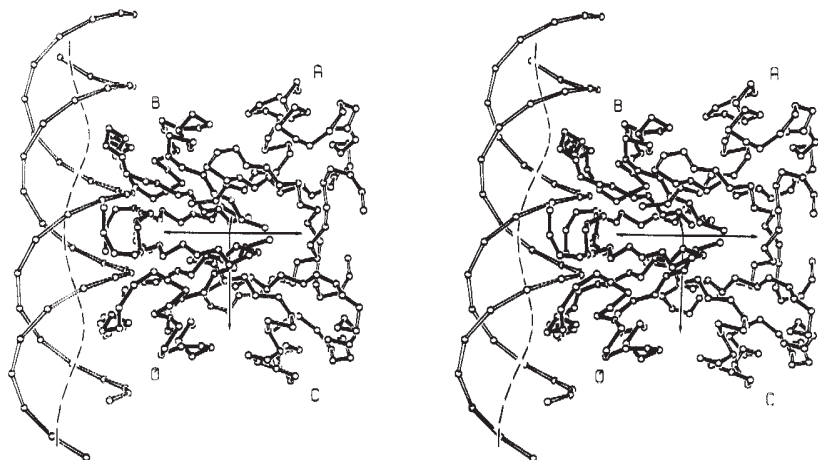


Fig. 7 Stereo view of the presumed binding of a tetramer of *cro* to DNA. Monomers O and B, seen interacting with the DNA, are drawn with solid bonds, whereas monomers A and C which form a second potential DNA binding region on the opposite side of the tetramer, are drawn with open bonds. The broken line follows the 'bottom' of the major groove of the DNA. The DNA and protein have been rotated 30° about the horizontal axis so that the direction of view is essentially along the major groove of the DNA.

against one side of the DNA, with most of the specific recognition occurring between side chains of the protein and the parts of the bases that are exposed in the major groove of the DNA. There is no apparent need to expose the bases further to facilitate recognition.

We thank Drs Howard Schachman and Harrison Echols for their help in instigating this project. We also thank our

colleagues, in particular D. Bacon, R. G. Fisher, I. Herskowitz, M. F. Schmid, G. R. Smith, L. F. Ten Eyck and L. H. Weaver, for assistance and advice. This work was supported in part by grants from the NSF (PCM-8014311), the NIH (GM 20066; GM 28138), the MRC of Canada and the M. J. Murdock Charitable Trust. D.H.O. was supported by a Damon Runyan-Walter Winchell Cancer Fund Postdoctoral Fellowship (DRG-303-F).

Received 10 December 1980; accepted 18 February 1981.

- Echols, H. A. *Rev. Genet.* **6**, 157-190 (1972).
- Herskowitz, I. A. *Rev. Genet.* **7**, 289-323 (1973).
- Ptashne, M. *et al. Science* **194**, 156-161 (1976).
- Ptashne, M. *et al. Cell* **19**, 1-11 (1980).
- Johnson, A., Meyer, B. J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1783-1787 (1978).
- Takeda, Y. *J. molec. Biol.* **127**, 177-191 (1979).
- Folkmanis, A., Takeda, Y., Simuth, J., Gussin, G. & Echols, H. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2249-2253 (1976).
- Takeda, Y., Folkmanis, A. & Echols, H. *J. biol. Chem.* **252**, 6177-6183 (1977).
- Hsiang, M. W., Cole, R. D., Takeda, Y. & Echols, H. *Nature* **270**, 275-277 (1977).
- Roberts, T. M., Shimatake, H., Brady, C. & Rosenberg, M. *Nature* **270**, 274-275 (1977).
- Green, D. W., Ingram, V. M. & Perutz, M. F. *Proc. R. Soc. A* **225**, 287-307 (1954).
- Matthews, B. W. in *The Proteins* Vol. 3. 3rd edn (eds Neurath, H. & Hill, R. L.) 403-590 (Academic, New York, 1977).
- Blundell, T. L. & Johnson, L. N. *Protein Crystallography* (Academic, London, 1976).
- Anderson, W. F., Takeda, Y., Echols, H. & Matthews, B. W. *J. molec. Biol.* **130**, 507-510 (1979).
- Anderson, W. F., Takeda, Y., Ohlendorf, D. H. & Matthews, B. W. *Proc. 7th Katzir-Katchalsky Conf. on Structural Aspects of Recognition and Assembly in Biological Macromolecules* March 1980 (in the press).
- Schmid, M. F. *et al. Acta crystallogr.* (in the press).
- Molnar, C. E., Barry, C. D. & Rosenberger, F. U. *Tech. Memo. No. 229* (Computer Systems Laboratory, Washington University, St. Louis, 1976).
- Blow, D. M. & Crick, F. H. C. *Acta crystallogr.* **12**, 794-802 (1959).
- Bricogne, G. *Acta crystallogr.* **A30**, 395-405 (1974).
- Richards, F. M. *J. molec. Biol.* **37**, 225-230 (1968).
- Colman, P. M., Jansonius, J. N. & Matthews, B. W. *J. molec. Biol.* **70**, 701-724 (1972).
- Maniatis, T. *et al. Cell* **5**, 109-113 (1975).
- Flashman, S. M. *Molec. gen. Genet.* **166**, 61-73 (1978).
- Meyer, B. J., Maurer, R. & Ptashne, M. *J. molec. Biol.* **139**, 163-194 (1980).
- Watson, J. D. & Crick, F. H. C. *Nature* **171**, 737 (1953).
- Arnott, S. & Hukins, D. W. L. *Biochem. biophys. Res. Commun.* **47**, 1504-1509 (1972).
- Church, G. M., Sussman, J. L. & Kim, S.-H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1458-1462 (1977).
- Johnson, A., Meyer, B. J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5061-5065 (1979).
- Sung, M. T. & Dixon, G. H. *Proc. natn. Acad. Sci. U.S.A.* **67**, 1616-1623 (1970).
- Adler, K. *et al. Nature* **237**, 322-327 (1972).
- Warrant, R. W. & Kim, S. H. *Nature* **271**, 130-135 (1978).
- McKay, D. & Steitz, T. A. *Nature* (this issue).
- Herskowitz, I. & Hagen, D. A. *Rev. Genet.* **14**, 399-445 (1980).
- Echols, H. in *The Molecular Genetics of Development* (eds Loomis, W. & Leighton, T.) 1-16 (Academic, New York, 1980).
- Johnson, A. thesis, Harvard Univ. (1980); and manuscript in preparation.

LETTERS

Solar quiescent prominences at 10.7 GHz

R. N. Bracewell

Electrical Engineering Department, Stanford University, Stanford, California 94305, USA

W. Graf

SRI International, Menlo Park, California 94025, USA

Filaments on the Sun, which are dark linear features in the light of the Fraunhofer $H\alpha$ line, are a mysterious feature of the solar surface. They are blade-like extensions of the chromosphere projecting to a height of 40,000 km into the corona, and may extend for up to 200,000 km along the surface. The fact that they may persist for a year or so, disappear abruptly and then reappear in the same positions shows that they are superficial indicators of deeper phenomena. There have been several reports¹⁻⁴ of reduced brightness in the radio spectrum associated with the $H\alpha$ filaments, leading to the widely accepted view that filaments are associated with broad dark radio features. This picture is, however, contradicted by our observations at 10.7 GHz of one solar filament during the solar eclipse of 12 October 1977.

Although the filaments are very thin, subtending about 8 geocentric arcs when seen vertically, nevertheless, "regions of reduced radio brightness situated above filaments" were detected¹ in 1959 and explained in terms of "relatively cold prominences which are optically thick to 8-mm radiation". Later attempts to discern filaments at radio wavelengths succeeded due to use of the 11 m mm-wave telescope at Kitt Peak and the 42 m telescope at Green Bank²⁻⁴. In all cases temperature depressions in the neighbourhood of filaments were reported. Khangil'din's¹ interpretation was absorption in the filamentary material; Drago and Felli⁴ suggested that the temperature defect arose from lack of coronal emission from the volume occupied by the filament. The width of the depression was confirmed at ~2 arc min by Straka *et al.*⁵ who proposed a model where the filament would exhibit a greater brightness temperature because the enhanced density within the filament would raise the height of the emission region. The general depression previously reported would then be due to the low-density 'cavity' surrounding the filament. Observations⁶ with the 100-m telescope at Effelsberg again confirmed the temperature depression. The depression was explained in terms of 'radio filament' for the 'temperature depressions' or 'absorption features' even though the width was perhaps five times larger than that of the dark optical filament. This interpretation in terms of absorption disagreed with Straka *et al.*'s model⁵ of the cavity concept because Kundu *et al.*⁶ believed that the cavity model

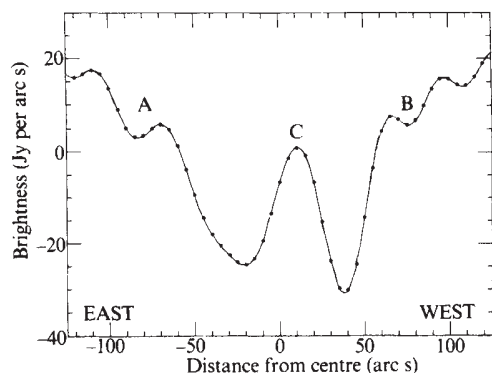


Fig. 1 One-dimensional brightness distribution across a temperature depression AB embracing a bright feature C (on a dark background) marking the position of a solar filament. The peak at C agrees in position with the dark H α filament photographed the same day at Big Bear Solar Observatory. The zero of the ordinate scale is at our standard quiet-Sun level. The brightness at the edges of the figure is slightly greater than zero because areas brighter than the quiet-Sun level fall within the fan beam of the antenna.

was inconsistent with the observations of prominences in emission at the limb^{7,8}. A coronal cavity alone was not considered sufficient by Raoult *et al.*⁹. Rao and Kundu¹⁰ used the Westerbork interferometer to study the solar limb at 6 cm and found a negative trough near the limb that was in the same place as a filament. After careful correction for grating responses they confirmed the Kundu interpretation.

On 12 October 1977 the Stanford five-element fast interferometer¹¹, which had a beamwidth of 17 arc s at 2.8 cm, was used to observe a quiescent filament during a solar eclipse. For

scan (Fig. 1) across the narrow dimension of the filament shown in Fig. 2. Other high-resolution solar observations with the Stanford instrument are discussed elsewhere¹².

The existence of a radio depression, postulated by Straka *et al.* to be associated with the cavity, is confirmed by the region of reduced brightness between A and B and superimposed on the depression is a very narrow emission feature C at the position of the filament itself. We also have several other observations of filaments most of which show bright features within the dark cavity. This emission feature in the cavity means that the objection against the model of Straka *et al.* needs revising.

It seems that the angular resolution of the 100-m Effelsberg telescope, while ranging from 5 down to 0.67 arc min over the wavelengths available, was not sufficient to show the very narrow bright feature. Likewise Rao and Kundu, although working with a synthesized beam of potentially adequate angular resolution, were limited to studying a polar filament at oblique incidence. In view of the present high-resolution observation obtained at a favourable position angle the concept of a broad low-brightness 'radio filament' should be replaced by a model of a narrow filament-associated emission feature embedded in a dark field that is associated with the cavity. The term radio filament should then be redefined as the radioemissive feature coincident with the optical filament.

The Drago and Felli calculation must now be revised to explain a bright feature instead of a dark one. We conclude that the deficit of coronal emission from the volume occupied by the filament must be compensated by extra emission from the filament, which therefore cannot be assumed to resemble undisturbed chromosphere but must have a higher brightness temperature.

The work was supported by the NSF Division of Atmospheric Sciences.

Received 13 August 1980; accepted 2 March 1981.

1. Khangil'din, U. V. *Soviet Astr. A-J* **8**, 234 (1964).
2. Buhl, D. & Tlamicha, A. *Astr. Astrophys.* **5**, 102 (1970).
3. Kundu, M. R. *Solar Phys.* **13**, 348 (1970).
4. Drago, F. C. & Felli, M. *Solar Phys.* **14**, 171 (1970).
5. Straka, R. M., Papagiannis, M. D. & Kogut, J. A. *Solar Phys.* **45**, 131 (1975).
6. Kundu, M. R., Fürst, E., Hirth, W., & Butz, M. *Astr. Astrophys.* **62**, 431 (1978).
7. Kundu, M. R. *Solar Phys.* **25**, 108 (1972).
8. Bell, M. B. *Solar Phys.* **27**, 137 (1972).
9. Raoult, A., Lantos, P. & Fürst, E. in *IAU Colloq. No. 44*, 131 (1978).
10. Rao, A. P. & Kundu, M. R. *Astr. Astrophys.* **86**, 373 (1980).
11. Bracewell, R. N. *et al. Proc. IEEE* **61**, 1249 (1973).
12. Falchi, A. D., Felli, M. & Tofani, G. *Solar Phys.* **48**, 59 (1976).



Fig. 2 H α spectroheliogram for 21 h UT, 12 October 1977 (courtesy of Big Bear Solar Observatory). The Moon's limb is seen receding from the Sun in the south-east quadrant, having just uncovered the filament under study. The position angle of the fan beam was 24° west of north. The north point of the Sun is at the top.

comparison, the 11-m telescope of the National Radio Astronomy Observatory has beamwidths of 3.5, 1.4 and 1.2 arc min at wavelengths of 9, 3.5 and 1.2 mm. The aim was to synthesize a microwave spectroheliogram, the east-west resolution being provided by the interferometer, and the north-south resolution by the motion of the Moon's limb. Favourable orientation of the filament enabled us to dispense with the Moon. We have synthesized a standard east-west brightness

Did ice streams carve martian outflow channels?

Baerbel K. Lucchitta

US Geological Survey, 2255 North Gemini Drive, Flagstaff, Arizona 86001, USA

Duwayne M. Anderson & Hitoshi Shoji

State University of New York at Buffalo, Buffalo, New York 14260, USA

Outflow channels on Mars¹ are long sinuous linear depressions that occur mostly in the equatorial area ($\pm 30^\circ$ lat.). They differ from small valley networks² by being larger and arising full born from chaotic terrains. Outflow channels resemble terrestrial stream beds, and their origin has generally been attributed to water³⁻⁵ in catastrophic floods^{6,7} or mudflows⁸. The catastrophic-flood hypothesis is derived primarily from the morphological similarities of martian outflow channels and features created by the catastrophic Spokane flood that formed the Washington scablands. These similarities have been documented extensively^{3,6,7}, but differences of scale remain a major problem: martian channel features are on the average much

larger than their proposed terrestrial analogues. We examine here the problem of channel origin from the perspective of erosional characteristics and the resultant landforms created by former and present-day ice streams and glaciers on Earth. From morphologic comparisons, an ice-stream origin seems equally well suited to explain the occurrences and form of the outflow channels on Mars, and in contrast with the hydraulic hypothesis, ice streams and ice sheets produce terrestrial features of the same scale as those observed on Mars.

The most prominent outflow channels originate in the chaotic terrains of Lunae Planum and Margaritifer Sinus, trend north across highland terrain, and debouch into the northern plains of Chryse Planitia. Many of the morphological characteristics of these channels correspond to those of terrestrial analogues produced by glacial action. For example, martian outflow channels widen, constrict, and anastomose around islands, and become deeply incised in places; ice streams in Antarctica anastomose around ice rises. Incised valleys are typical of glacial scenes in Yosemite Park in California and in fjords in northern Canada, Greenland, Scandinavia, and Antarctica. Hanging valleys occur along martian outflow channels where distributaries and tributaries merge with the main channel. Hanging valleys are a classical landform of terrestrial glacial regions. Where martian channels constrict, they are observed to have U-shaped crossprofiles (Fig. 1); this geomorphologic shape, on Earth, is characteristic of glacial erosion. Figure 1 also clearly shows long even scour marks on martian valley walls and similar scour marks consisting of glacial terraces in the Black and Oshetna Valleys in Alaska. Longitudinal ridges and grooves are conspicuous on martian valley floors, where they have been interpreted as being caused by roller vortices⁹. Such longitudinal grooves are also characteristic of terrestrial glaciated landscapes, where they are developed in moraines or bedrock. They abounded under the former Laurentide ice sheet, and are clearly visible on Landsat images of the Northwest Territories of Canada (Fig. 2) and in large glaciated valleys in Alaska and Iceland. In analogy with glacial grooves on Earth, the grooves on Mars could indicate that ground moraines are

present in places. (End moraines, however, are not to be expected unless steady ice supplies and melt locations were maintained for extended times.) Although each of the individual landforms on Mars could have been formed by other processes, the large number of landforms pointing towards glacial action is highly suggestive of an ice origin.

The streamlining of islands on Mars has been cited as a critical indication of catastrophic flooding because they resemble the streamline forms visible in the scablands of Washington State^{7,10}. We have compared the martian islands with ice rises in Antarctica, which are presently being sculptured by active ice streams (Fig. 3), and have plotted their dimensions on Baker's graphs. Figure 4 shows that the Antarctic ice rises lie on the same trendline as the scabland and martian islands suggesting that ice sculpturing is at least as likely a process responsible for the formation and streamlining of the martian islands as catastrophic flooding. We have also compared the elongation (length-to-width ratio) of martian islands with those of ice rises. Baker and Kochel's¹⁰ calculated average elongation of 2.7 for Kasei Vallis islands agrees exactly with the value we have calculated for 26 islands near the mouth of Tiu Vallis. Baker and Kochel found that small scabland islands (1 km long) have elongations of 3.0 and large islands (100-km long) elongations of 4.1. Thus, scabland islands diverge somewhat from the elongation of martian islands as they approach the martian features in size. In contrast, Antarctic ice rises, which are similar in size to the largest martian islands, have an average elongation of 2.5. This value agrees closely with the values derived from dimensions of the martian islands.

The dimensions of fjords and ice streams in Antarctica are also comparable with those of martian outflow channels, even though these channels are very long, wide, and locally deep, and dwarf many terrestrial rivers and alpine glacial valleys. The width of martian outflow channels varies widely; the maximum width of several channels ranges from 50–70 km to as much as 180 km across the grooved terrain in Kasei Vallis. Several outlet glaciers in Antarctica have maximum widths of near 50 km. The mouth of Lambert Glacier is as wide as 125 km, and we

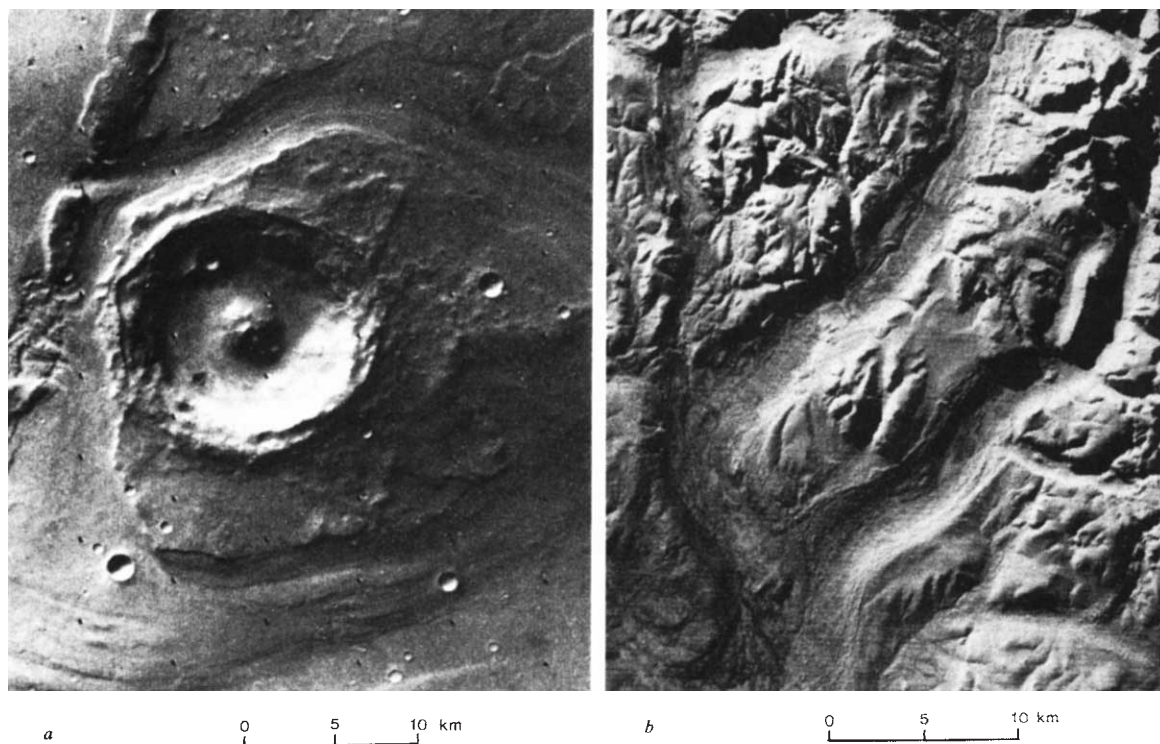


Fig. 1 U-shaped valleys on Mars and Earth. Where martian channels constrict, they may acquire U-shaped cross profiles, "similar to terrestrial glaciated valleys. Note similarity in long linear scour marks on valley walls, and valley curvature. *a*, Chryse Planitia. Breached wrinkle ridge on left, channel around crater in centre. Flow was towards right. North is towards the top. (Viking 20A62). *b*, Black and Oshetna glacial valleys, Talkeetna Mountains, Alaska. Glacial flow was towards top. North is towards the bottom (Landsat 1102-20452-5).

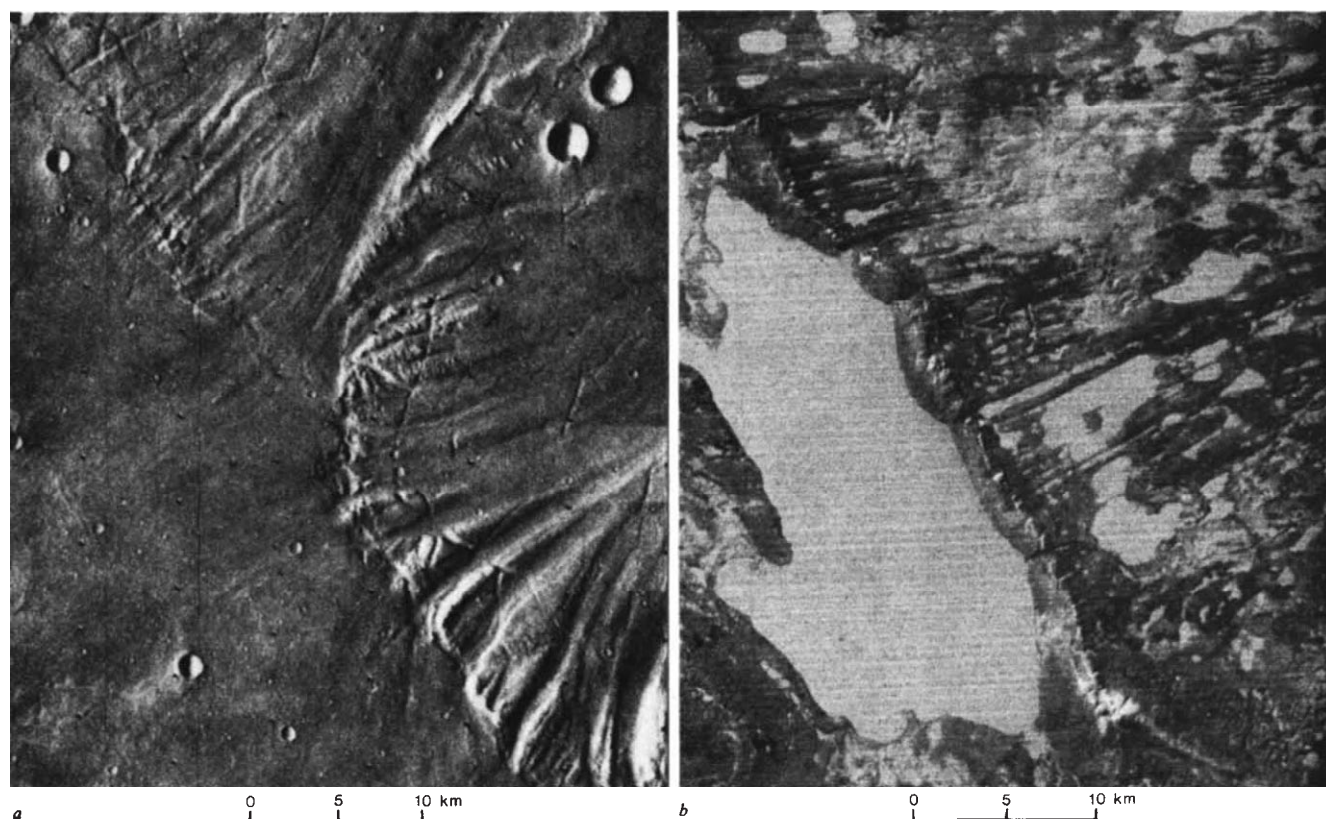


Fig. 2 Grooved terrain on Mars and Earth. Cross ridges in centre of images apparently dammed and ponded carving material, which eventually breached and deeply scoured the terrain at right. Flow was towards the right. *a*, Breached wrinkle ridge in Chryse Planitia and channel scour marks. North is towards the top. (Viking 46A65). *b*, Scarp near Lac Belot, Northwest Territories, Canada, and glacial grooves. North is towards the bottom. (Landsat 1594-19302-7).

measured widths of as much as 180 km across the grooved terrain of Laurentide ice sheets in the Northwest Territories of Canada. The comparable width of grooved terrain in Kasei Vallis and Canada suggests that some of the ice on Mars may have been in sheets rather than valley glaciers.

Scarp heights of the martian channels also compare favourably with terrestrial glacial scarps. Scarps flanking Kasei Vallis lie near 2,000 m, and those at the mouth of Tiu and Ares Valles near 600 m. Grooves on the level ground above and below some of the scarps indicate that the ice may locally have inundated the scarps and may have been thicker than 2,000 m. Such heights are not excessive for terrestrial glacial terrain; scarp heights on Lambert Glacier in Antarctica are as much as 3,000 m, and the glacier presently is thicker than 2,000 m (ref. 11). Ice streams into the Ronne Shelf near the Ellsworth Mountains are 1,000–2,000 m thick¹² (Fig. 3). Pleistocene ice streams in Iceland are deduced to have been 1,000 m thick (based on the height of tablemountains)¹³, and the present-day Antarctic and the former Laurentide shields are 3,000–4,000 m thick¹⁴.

To determine whether ice sheets would have moved in martian conditions, it is important to obtain information on ice-surface gradients. Because no ice sheets are presently in the areas of interest, only the gradients of the martian channel beds can be obtained. For this discussion, however, ice-surface gradients can be assumed to be parallel. Regional gradients are very low and range from $<0.1^\circ$ to 0.25° ; local segments are steeper. Low ground gradients are no deterrent to glacial movement on Earth, however. For instance, the Yenta River in Alaska has a gradient of 0.06° , and the Wilkes subglacial basin in Antarctica has a gradient of $\sim 0^\circ$ for nearly 2,000 km. Thus, ground gradients apparently are not a critical determinant for glacial motion; indeed, glaciers are known to move uphill under certain circumstances. Terrestrial ice-sheet surface gradients range from 0.10° to 0.25° , from the Antarctic highs to outlet

glaciers on the coast, and thus are comparable to the gradients of martian channel beds.

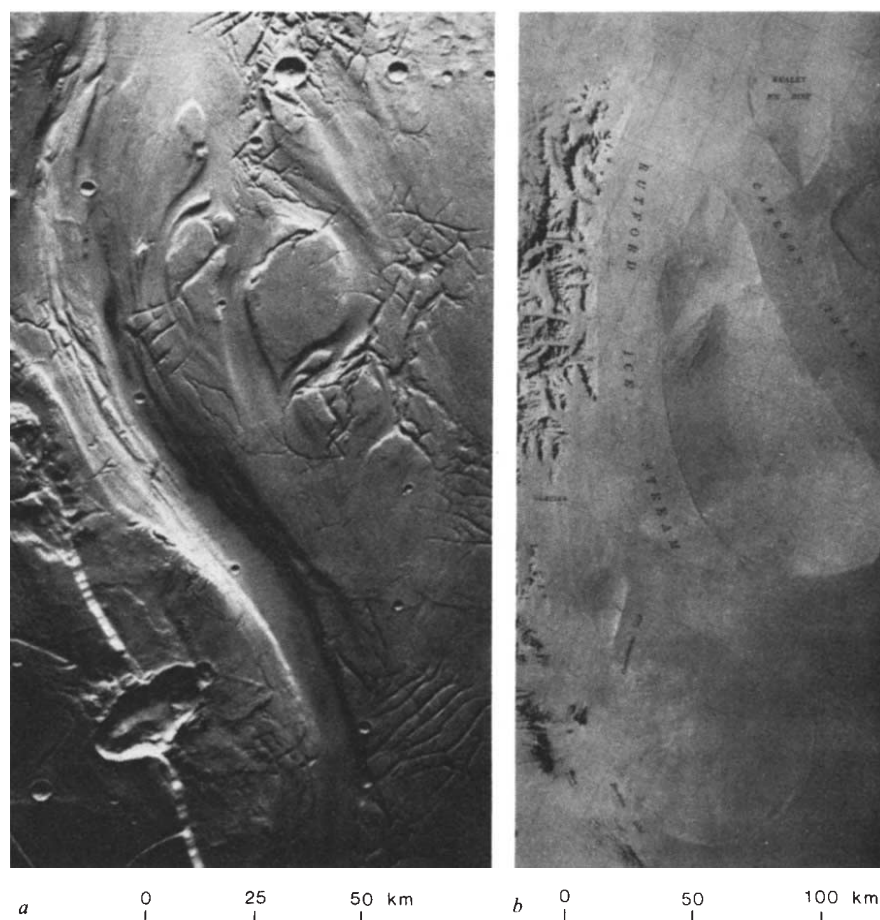
It remains to be considered whether or not ice could have moved through martian outflow channels, given present and past martian environmental conditions. Sharp¹⁵ argued against the likelihood of glacial motion on Mars; however, he only considered the cold polar areas and not the warmer equatorial belt, where most outflow channels occur. Existing data on ice-deformation processes have been collected and displayed by Shoji and Higashi¹⁶ in a plot of shear stress versus temperature. Taking nominal values of surface slope (0.1 – 0.25°) and ice thickness (1,500–3,000 m), basal shear stresses (τ) may be calculated from the relation

$$\tau = \rho g y \sin \alpha \quad (1)$$

where ρ , g , and y are the ice density, gravitational acceleration, and depth, respectively, and α is the surface slope. When these shear stresses are plotted against martian equatorial temperatures (0° to -60°C), the conventional terrestrial processes characteristic of ice-sheet deformation—namely, dislocation creep and lattice-diffusion creep—also seem to apply to ice deformation on Mars. In addition, the shear stress–temperature field that is defined for Mars includes all the measured shear stress–temperature data obtained at various depths in drill holes at Byrd Station, Antarctica, and Camp Century, Greenland. These results strikingly illustrate the similarity in conditions on the two planets, notwithstanding the generally lower temperatures now prevailing on Mars.

Nominal strain rates, given in the plot of shear stress versus temperature by Shoji and Higashi¹⁶, vary by many orders of magnitude and diminish with decreasing temperatures and shear stress. These data suggest that ice, in present-day martian conditions, would have moved only very slowly. Several other factors, however, could increase the strain rate. Preferred crystal

Fig. 3 Streamlined island on Mars and Earth. Flow diverging around obstacles formed by islands on Mars and ice rises on Earth. Martian islands tend to be smaller than ice rises. Flow is towards the bottom. *a*, Kasei Vallis. North is towards the right. (Viking 519A29). *b*, Ice rises in Ronne ice shelf, Antarctica. Ellsworth Mountains on left. Radio echo sounding suggests subglacial breaching of ice rises by fjordlike troughs, which would further enhance their similarity to martian islands. north is towards the top right. (Satellite image map, Ellsworth Mountains, Antarctica, US Geological Survey, original scale 1:500,000).



orientation and occluded gas bubbles in polycrystalline ice sheets act to increase strain rates at given temperatures and shear stresses by as much as 100-fold^{17,18}. Strain rates can also be increased once flow by dislocation creep is initiated by local steeper slopes and the ice mass undergoes frictional heating. The energy thus released internally by an ice sheet flowing at a velocity of 20 m yr^{-1} is roughly equal to the normal geothermal flux. An ice velocity of 20 m yr^{-1} is not extraordinary and, on

Earth or Mars, could be considerably larger if ice-surface slopes increased to several degrees. Artesian water under the ice could also have contributed to lubrication of the bases of martian ice masses and thus accelerated their motion and increased their erosive power. The pronounced grooving on the channel floors indeed suggests that the ice moved over the ground and eroded powerfully. A warmer climate in the past, of course, would have favoured faster glacial motion.

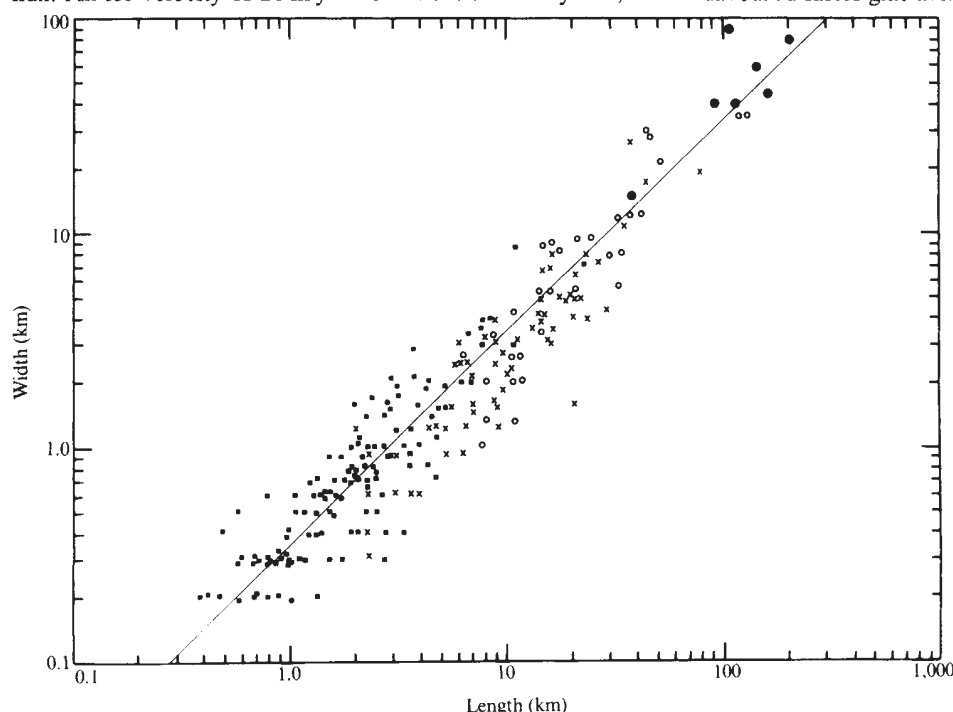


Fig. 4 Dimensions of streamlined islands. Plotted scabland flood islands (■) and martian islands are from ref. 7; ×, Maja Vallis; ○, Kasei Vallis. Plotting of Antarctic ice-rise dimensions (●) on same trendline as scabland and Mars values indicates that sculpturing by ice is as likely as catastrophic flooding for causing streamlining.

Ice may have accumulated in the source area of the martian channels from atmospheric precipitation, wind-drifted frost and dust deposits, and from the ground as springs. All these processes no doubt contributed, but springs must have been dominant because the chaotic terrain, the source area of the channels, is highly disturbed, apparently owing to the removal of large quantities of material. The groundwater may have been under artesian pressure as proposed by Carr¹⁹, and created ice domes that then flowed to adjacent depressions to carve the channels. Alternatively, the water could have formed ice covers over incipient rivers that required only normal discharges²⁰. The river ice may have consolidated into ice drives⁷ and, eventually, ice streams. The internal deformation and flow of such ice streams may then have imparted to the martian channels their distinctive and spectacular morphology.

We thank H. Masursky, M. Grolier and V. Baker for helpful reviews (although their point of view is not necessarily that of the authors). This work was supported by NASA contract W14,575 and grant 150-6330A.

Received 22 December 1980; accepted 8 March 1981.

1. Sharp, R. P. & Malin, M. C. *Bull. geol. Soc. Am.* **86**, 593–609 (1975).
2. Pieri, D. C. *Science* **210**, 895–897 (1980).
3. Baker, V. R. & Milton, D. J. *Icarus* **23**, 27–41 (1974).
4. McCauley, J. F. et al. *Icarus* **17**, 289–327 (1972).
5. Masursky, H., Boyce, J. M., Dial, A. L., Schaber, G. G. & Strobell, M. E. *J. geophys. Res.* **82**, 4016–4038 (1977).
6. Baker, V. R. *Science* **202**, 1249–1256 (1978).
7. Baker, V. R. *J. geophys. Res.* **84**, 7985–7993 (1979).
8. Nummedal, D. in *NASA Tech. Mem.* **79729**, 257–259 (1978); **81776**, 289–291 (1980).
9. Thompson, D. E. *Geophys. Res. Lett.* **6**, 735–738 (1979).
10. Baker, V. R. & Kochel, R. C. *Proc. 9th Lunar planet. Sci. Conf.* 3193–3203 (1978).
11. Morgan, V. I. & Budd, W. F. *J. Geol.* **15**, 103–111 (1975).
12. Swinbank, C. *Phil. Trans. Soc. B* **279**, 161–183 (1977).
13. Preusser, H. *Landscapes of Iceland; Types and Regions* (Junk, The Hague, 1976).
14. Sugden, D. E. *Arctic Alp. Res.* **9**, 21–47 (1977).
15. Sharp, R. P. *J. Glaciol.* **13**, 173–185 (1974).
16. Shoji, H. & Higashi, A. *J. Glaciol.* **21**, 419–427 (1978).
17. Garfield, D. E. & Ueda, H. T. *J. Glaciol.* **17**, 29–34 (1976).
18. Whillans, I. M. *J. Glaciol.* **18**, 359–372 (1977).
19. Carr, M. H. *J. geophys. Res.* **84**, 2995–3007 (1979).
20. Wallace, D. & Sagan, C. *Icarus* **39**, 385–400 (1979).

Decade fluctuations in geomagnetic westward drift and Earth rotation

Jean-Louis Le Mouél*, Theodore R. Madden†, Joël Ducruix* & Vincent Courtillot‡*

* Institut de Physique du Globe, Université Paris VI, 75230 Paris Cedex 05, France

† Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡ UER des Sciences Physiques de la Terre, Université Paris VII 75230 Paris Cedex 05, France

Numerous studies of the variations of the Earth's rotation have been made since 1900 (summarized in refs 1, 2). It seems well established that variations in the rotation with periods equal to or less than 2 yr are due to variations in the zonal circulation of the atmosphere^{2–4}. However, there is no general agreement about the cause of the longer-period variations of the Earth's rotation, sometimes known as the 'decade variations', although the most favoured explanation is electromagnetic core–mantle coupling (see ref. 5). Several studies^{6–8} of the correlation between Earth rotation and magnetic field variations have obtained marginally positive results. We have been able to clarify such a correlation over the past 120 yr by studying those magnetic field indicators that are the most sensitive to variations in westward drift of the outer regions of the core. We suggest here that the correlation can be extended back ~300 yr. Simple models of core–mantle coupling, based on Bullard's ideas, give reasonable quantitative agreement with the observations.

In 1951, Stoiko⁹ compared the variations of the Earth's magnetic field intensity observed at Paris with the variations of

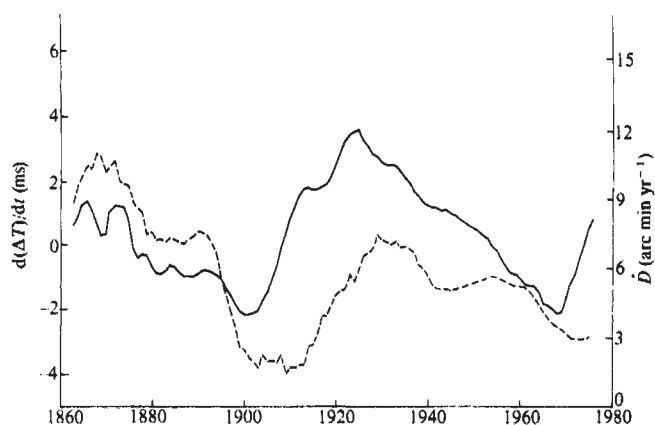


Fig. 1 Plot of the secular variation of declination $\dot{D}$ (solid curve and right hand scale) and of the excess length of day $d(\Delta T)/dt$ (dashed curve and left-hand scale) for the period 1861–1978. ΔT data are from refs. 18, 19. The relative changes in the Earth rotation rate $\Delta\Omega/\Omega$ are obtained from the $d(\Delta T)/dt$ values through multiplication by -1.157×10^{-8} .

the length of day for the period 1883–1948 and claimed a good correlation between the two. In 1953 Vestine⁶ proposed a correlation between the westward drift of the eccentric dipole (the magnetic centre C) and the variations of the Earth's rotation rate Ω for the period 1885–1945, and gave a qualitative interpretation of this result using Bullard's model. Vestine and Kahle¹⁰ extended the correlation up to 1965. A puzzling feature of this correlation is that the magnetic variations seem to lag behind the Ω variations. The variations of the longitude of C for the period 1945–75 have been found¹¹ to have a poor correlation with Ω . More recently, it has been shown^{12,13} that, starting in 1969, there was a rapid acceleration of the secular variation of the east (Y) component of the magnetic field across large regions of the Northern Hemisphere and that a similar event occurred around 1900 (although it is more difficult to ascertain its global nature because of the lack of magnetic observatories at that time). Minima in Ω were also noted at about the same times. It is this correlation between the Y variations and Ω which we consider and extend here.

Motions of the top layers of the core are responsible for most magnetic variations seen at the Earth's surface on the decade time scale; the westward drift of magnetic features which has long been noted^{14,15} can be taken as evidence for a westward drift of the upper core. Hide¹⁶ has pointed out that such motions could be the result of slow MHD–Rossby waves, but the simultaneity of the changes in these motions seen around the world makes this possibility unlikely.

To study the westward drift of magnetic features we have to distinguish between the drift effect and other causes of magnetic variations¹⁷. The magnetic secular variation can be written as:

$$\frac{\delta B}{\delta t} = \frac{\partial B}{\partial t} + U_{\phi} \frac{\partial B}{\partial \phi}$$

where $\delta B/\delta t$ is the magnetic variation seen at magnetic observatories; the $\partial B/\partial t$ term accounts for the effects of non uniform and north–south motions (as well as the less important effects of magnetic diffusion), U_{ϕ} is the magnitude of the westward drift as seen at the Earth's surface and $\partial B/\partial \phi$ is the longitudinal gradient of the magnetic field as seen at the surface.

Declination D (or the closely associated east component Y) is especially suited to studies of zonal motions because (1) it has weak gradients in the north–south directions compared with the other components and (2) it was more accurately and frequently determined in earlier times.

To study the detailed time history of the magnetic secular variations data from continuously recording magnetic observatories are needed. Most of the early magnetic observatories are located in Europe and it is fortunate that here the $\partial Y/\partial \phi$ term is large and smoothly varying. Thus $\delta Y/\delta t$ is dominated, at

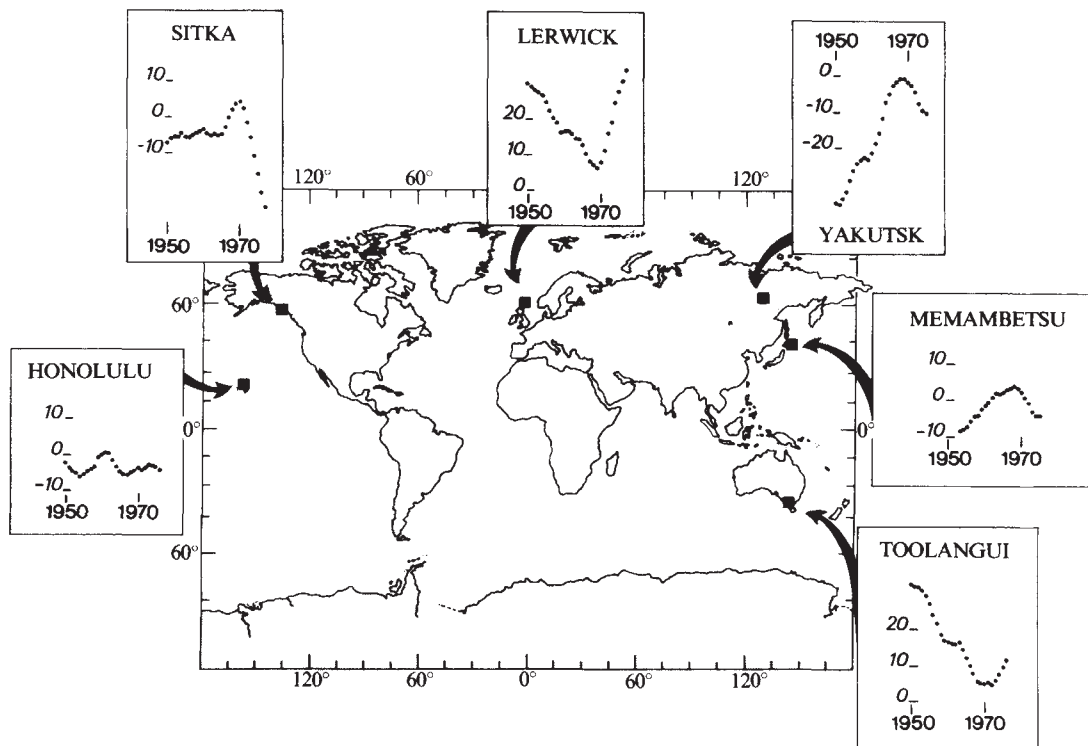


Fig. 2 Selected $\delta Y/\delta t$ variations (in nT yr^{-1}) around the period of apparent westward drift minimum of 1969.

least for the past century, by the term $U_\phi(\partial Y/\partial \phi)$, and $\dot{Y}$ (or $\dot{D}$) should be proportional to U_ϕ . For Paris $\partial Y/\partial \phi$ is $\sim 150 \text{ nT per deg}$ and $\partial D/\partial \phi$ is $\sim 0.4\text{--}0.5$. If core–mantle coupling is the cause of the decade variations in Earth rotation one should expect a good correlation between $\dot{Y}$ (or $\dot{D}$) as seen in Europe and Ω .

Figure 1 displays $\dot{D}$ and $\Delta\Omega/\Omega$ for the period 1861–1978. The D values are from a compilation of available data from the Paris area. These data can be considered as representative of the western European area, as has been shown for the recent observatory data¹³. $\dot{D}$ has been averaged over 5 yr. The Ω data for 1861–1978 are from refs 18, 19.

The data of Fig. 1 show a good correlation between $\dot{D}$ and Ω , significantly better than in any previous study (see Fig. 4 of ref. 10 or Fig. 4 of ref. 8). The magnetic variations lead, rather than lag⁷, the Ω variations by $\sim 10 \text{ yr}$. (The coherency peaks fairly sharply at about 0.8 for this value of the lag. A more sophisticated correlation analysis is not warranted for the kind of data shown in Fig. 1 and the eye is still the best judge.) These variations are quite pronounced. From $\dot{D}$ one can infer that the westward drift decreased from ~ 0.3 to 0.1 deg yr^{-1} between 1870 and 1900, increased again from ~ 0.1 to 0.4 deg yr^{-1} between 1900 and 1925 and decreased again to $\sim 0.1 \text{ deg yr}^{-1}$ in 1969 (see also ref. 11). Since then a rapid acceleration of the drift has been seen, which is quite similar to that observed just after 1900.

The most pronounced features of secular variation are the discontinuities in the slope of Y seen in 1900 and 1969, and somewhat less clearly in 1925, which we interpret as sudden changes in the acceleration of the westward drift, $\dot{U}_\phi$. The number and distribution of magnetic observatories in 1900 were too limited to ascertain the global character of the drift acceleration changes, but by 1969 the situation was greatly improved. Fig. 2 shows selected results from a preliminary study, still in progress, which shows that in most areas where we expect to see the effects of rapid changes in the slope of U_ϕ they are indeed observed. The examples in Fig. 2 come both from regions where $\partial Y/\partial \phi$ is large and smoothly varying (for example, Europe^{12,13}) and from regions where $\partial Y/\partial \phi$ is close to zero (for example, Honolulu): in the former a clear extremum of

$\partial Y/\partial \phi$ is observed, in the latter there is none. However, in some cases (for example, South America) discrepancies are found which may be due to the evolution of more localized magnetic features. In any case, the magnetic data are not incompatible with a global zonal circulation pattern existing in the upper core which underwent a rapid change of acceleration in 1969^{13,20}.

We can only attempt here to identify some of the parameters that are relevant in a quantitative study of core–mantle coupling and estimate their magnitudes. We follow Bullard's model of a convecting core which causes differential rotation between the outer regions and the inner regions. This differential rotation creates a toroidal magnetic field in the core which penetrates into the lower mantle¹⁵. Thus, the mantle, upper and inner core are all electromagnetically coupled, and, if the convection regime changes, the core and mantle will all undergo changes in their angular velocities, the net angular momentum remaining constant. When the upper core region is thin, the mantle follows the inner core and will accelerate when the upper core decelerates (see ref. 20, with lower mantle conductivity estimates from ref. 21): this is what is observed. The thickness of the upper core zone can be estimated by assuming that the mantle and lower core are rotating together. Using the law of conservation of momentum and the observed proportionality rate between $\delta\Omega$ and δU_ϕ (with $\delta\Omega \sim 3 \times 10^{-12} \text{ rad s}^{-1}$ and $\delta U_\phi \sim 0.2 \text{ deg yr}^{-1}$) a thickness of 100 km is found for this upper region. The time lag of the changes in Ω relative to U_ϕ is a measure of the magnitude of the coupling between the mantle and the inner core, which in turn depends on the magnitude of the toroidal field in the core and the conductivity of the lower mantle. The observed time lag can be accounted for by the above model (Le Mouél and Courtillot, in preparation).

We have also compared data for the period 1700–1870. Despite low reliability, these older data do not display the dramatic changes that have occurred since 1870. These observations seem to imply that the period 1700–1870 was a more stable regime for core motions than the following century. The similarity of the magnetic situation in 1900 and 1969 leads us to predict that the Earth rotation is presently at the beginning of a period of acceleration.

Received 1 December 1980; accepted 20 February 1981.

1. Munk, W. H. & McDonald, G. J. F. *The Rotation of the Earth* (Cambridge University Press, London 1960).
2. Lambeck, K. *The Earth's Variable Rotation. Geophysical Causes and Consequences* (Cambridge University Press, London, 1980).
3. Cazenave, A. thesis, Toulouse, Univ. (1975).
4. Lambeck, K. & Cazenave, A. *Phil. Trans. R. Soc. A* **284**, 495–506 (1977).
5. Runcorn, S. K., in *Earthquake Displacement Fields and the Rotation of the Earth* (eds. Mansinha, L. et al.) 181–187 (Reidel, Dordrecht, 1970).
6. Vestine, E. H. *J. geophys. Res.* **58**, 127–145 (1953).
7. Kahle, A. B., Ball, R. H. & Cain, J. C. *Nature* **223**, 165 (1969).
8. Yukutake, T. *J. Geomagn. Geoelectr.* **25**, 195–212 (1973).
9. Stoyko, N. C. *r. heb. Séanc. Acad. Sci., Paris* **205**, 1577–1579 (1937); **223**, 80–82 (1951); **224**, 1798–1800 (1952).
10. Vestine, E. H. & Kahle, A. B. *Geophys. J. R. astr. Soc.* **15**, 29–37 (1968).
11. Harwood, J. M. & Malin, S. R. C. *Nature* **259**, 469–471 (1976).
12. Courtillot, V., Ducruix, J. & Le Mouél, J. L. C. *r. heb. Séanc. Acad. Sci. Paris* **D287**, 1095–1098 (1978); **B289**, 173–175 (1979).
13. Ducruix, J., Courtillot, V. & Le Mouél, J. L. *Geophys. J. R. astr. Soc.* **61**, 73–94 (1980).
14. Halley, E. *Phil. Trans. R. Soc.* **12**, 208–221 (1863).
15. Bullard, E. C., Freedman, C., Gellman, H. & Nixon, J. *Phil. Trans. R. Soc. A* **243**, 67–92 (1950).
16. Hide, R. *Phil. Trans. R. Soc. A* **259**, 615–650 (1966).
17. Yukutake, T. *Bull. Earthq. Res. Instr. Tokyo* **40**, 1–64 (1962).
18. Morrison, L. V. *Geophys. J. R. astr. Soc.* **58**, 349–360 (1979).
19. Guinot, B. *La Terre, les Eaux, l'Atmosphère* Ch. 2 (Rotation de la Terre, Bureau des Longitudes, Paris, 1977).
20. Le Mouél, J. L. & Courtillot, V. *Phys. Earth planet. Inter.* (in the press).
21. Achache, J., Courtillot, V., Ducruix, J. & Le Mouél, J. L. *Phys. Earth planet. Inter.* **23**, 72–75 (1980); *Geophys. J. R. astr. Soc.* (in the press).

Phonon scattering at grain boundaries in heavily doped fine-grained silicon-germanium alloys

D. M. Rowe, V. S. Shukla & N. Savvides*

Department of Physics, Electronics and Electrical Engineering, UWIST, Cardiff CF1 3NU, UK

Thermocouples made from heavily doped n- and p-type silicon-germanium alloys are used in the nuclear powered thermoelectric generators which provides onboard power to the American Voyager spacecraft. The isotopic heat source is expensive and any improvement in the generators' heat energy to electrical energy conversion efficiency would save both money and weight. The possibility of reducing the thermal conductivity of thermoelectric semiconductor alloys by compacting fine grained material has been considered and various estimates have been made of the reduction in thermal conductivity with grain size. The electrical properties of hot pressed compacts of heavily doped silicon germanium alloys which possess density values close to that of single crystal are unaffected by grain boundary scattering effects. Consequently, this method of preparation holds out a possibility of improving the 'figure of merit' of the thermocouple material and hence the conversion efficiency of the generator. Here the first measurements of the thermal diffusivity of fine grained hot pressed compacts of heavily doped n-type silicon germanium alloy are reported. The compacts investigated possessed grain sizes (L) in the range $10 < L < 25 \mu\text{m}$, $5 < L < 10 \mu\text{m}$ and $< 5 \mu\text{m}$. The results confirm that the lattice thermal conductivity decreases with a reduction in grain size. In the compact with $L < 5 \mu\text{m}$, this reduction is ~28% compared with 'single crystal' material. Within experimental error (<5%) the Seebeck coefficient and electrical resistivity do not change with grain size. This implies that the conversion efficiency of a thermoelectric generator employing silicon-germanium thermocouples would be significantly improved through the use of fine-grained thermocouple material.

The conversion efficiency of a semiconductor thermocouple depends on the temperature difference over which the device operates, its average temperature and the so-called figure of merit Z for the thermocouple material. $Z = S^2\sigma/K$ where K is the thermal conductivity, S the Seebeck coefficient and σ the

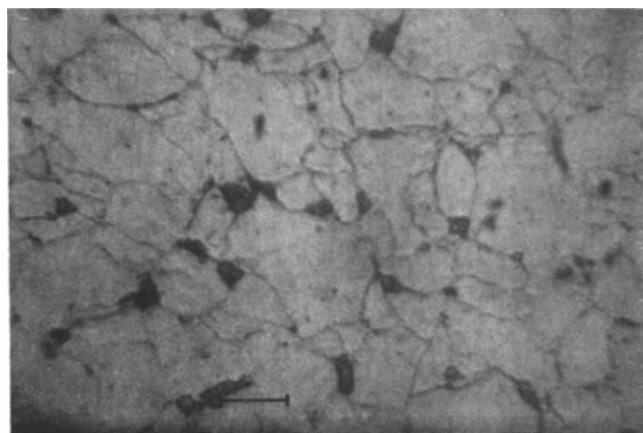


Fig. 1 Microphotograph of a typical polished surface of a compact after etching with CP4. Grain size is in the range $10 < L < 25 \mu\text{m}$. Scale bar, $14 \mu\text{m}$.

electrical conductivity. The thermal conductivity is the sum of the contribution arising from the lattice waves (phonons) K_L and a contribution K_e from the carriers present. All three parameters in the figure of merit vary with the carrier concentration and Z is optimized at around 10^{26}m^{-3} . At this high carrier concentration K_L still accounts for more than 75% of the thermal conductivity².

Interest has grown recently in the use of hot pressing techniques. Apart from facilitating device fabrication, the method of preparation has held out a possibility of reducing the lattice thermal conductivity and hence improving the figure of merit without adversely affecting the other thermoelectric parameters—an effect attributed to phonon scattering at the grain boundaries³.

The role of boundary scattering in limiting the phonon mean free path has long been considered as a low temperature phenomenon, where in the absence of other phonon scattering mechanisms boundary scattering effects are observed as the phonon mean free path approaches the sample dimensions. In solid solutions such as silicon-germanium alloys, however, because a large proportion of the heat is carried by low frequency phonons, boundary scattering effects can manifest themselves in the region of the Debye temperature and above and in principle lead to a reduction in the lattice thermal conductivity.

In thin films of germanium the carrier mean free path is unaffected by boundary scattering down to thicknesses of $0.1 \mu\text{m}$ (ref. 4) and measurements on high density heavily doped hot pressed silicon germanium alloys with grain sized $< 70 \mu\text{m}$ (ref. 5) and $< 10 \mu\text{m}$ (D.M.R. and N. Madini, unpublished data) indicate that the electrical conductivity and Seebeck coefficient differ from single crystal values by less than 5%.

In 1976 a theoretical model was proposed for heavily doped fine grained silicon germanium alloys⁶ but the combination of hot pressing and heavy doping complicates the situation, and in spite of refinements to the model there had been no satisfactory quantitative description of the thermal and electronic behaviour⁷. However, some theoreticians agree that in alloys with a grain size of $< 1 \mu\text{m}$ a reduction in K_L of 20–30% compared with the K_L for single crystal material would be observed.

Table 1 Room temperature properties of investigated Si-Ge alloys

Specimen	K ($\text{W m}^{-1} \text{K}^{-1}$)	S ($\mu\text{V K}^{-1}$)	$\rho = \sigma^{-1}$ ($\times 10^{-5} \Omega\text{m}$)	K_e ($\text{W m}^{-1} \text{K}^{-1}$)	K_L ($\text{W m}^{-1} \text{K}^{-1}$)
Single crystal	5.07	116	0.85	0.76	4.31
$10 < L < 25 \mu\text{m}$	4.93	129	1.05	0.72	4.21
$5 < L < 10 \mu\text{m}$	4.26	133	1.10	0.69	3.57
$L < 5 \mu\text{m}$	3.82	131	1.05	0.72	3.10

* Present address: Division of Chemistry, National Research Council, Ottawa, Canada K1A 0R9.

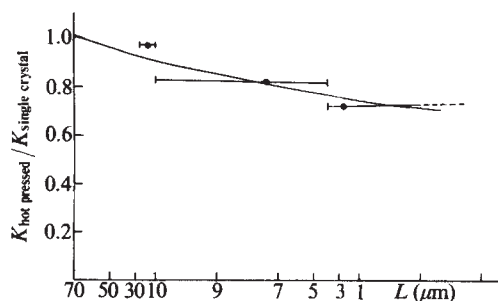


Fig. 2 $K_{\text{hot pressed}}/K_{\text{single crystal}}$ as a function of grain size for heavily-doped silicon-germanium alloys.

We investigated material cut from a single ingot of homogeneous, zone levelled $\text{Si}_{63.5}\text{Ge}_{36.5}$ alloy doped n-type with phosphorus to about $1 \times 10^{26} \text{ m}^{-3}$. A disk shaped sample with suitable dimensions for thermal diffusivity measurement (6 mm diameter $\times$ 1 mm thick) was cut from a slice from the ingot using an ultrasonic cutter. The remainder of the slice was crushed in an agate pestle then ground in an agate vibromill. The fine powder was sieved through a set of British Standard microplate sieves assisted by ultrasonic vibration⁸. Compacts 10 mm in diameter and 3 mm thick were prepared from starting powders possessing a grain size in the range $10 < L < 25 \mu\text{m}$, $5 < L < 10 \mu\text{m}$ and $< 5 \mu\text{m}$ using a vacuum hot pressing technique⁹. The density of the compacts was better than 98% of that of the single crystal starting material. A typical microphotograph of the polished surface of a compact is shown in Fig. 1, and it is apparent that little, if any, grain growth has occurred during compaction. A disk shaped sample was cut from each of the compacts and all four disks annealed at 1,000 °C under vacuum for 1 h followed by quenching in cold water to normalize the carrier concentration and hence the value of K_e (ref. 10).

The thermal diffusivity α of the 'single crystal' starting material and of the three compacts was measured using a system based on the laser flash technique¹¹. Measurements on ARMCO iron indicated that the accuracy of the system was better than $\pm 2\%$ at room temperature. Reproducibility of measurements was about the same. Seebeck coefficient data were obtained using a hot probe method (accuracy $\pm 3\%$) and carrier concentration and electrical resistivity values derived from carrier concentration versus Seebeck coefficient¹² and electrical resistivity curves¹³.

The experimental results are displayed in Table 1. Thermal conductivity values were obtained from the measured thermal diffusivity using the relationship $K = C_p \alpha$ where C_p is the specific heat per unit volume at constant pressure and values appropriate to the alloy investigated are taken from the literature¹². The electronic contribution to the thermal conductivity is given by $K_e = \mathcal{L}(k/e)^2 \sigma T$. Where $\mathcal{L}$ depends on the scattering mechanisms of the charge carriers and for the material investigated is ~ 2.9 (ref. 7). The Seebeck coefficient of the three compacts are very similar but all are higher than the 'single crystal' starting material. This increase in Seebeck coefficient and the corresponding increase in electrical resistivity is due to a relatively small loss of dopant during the hot pressing operation. The lattice thermal conductivity decreases with a reduction in grain size and for the compact with a grain size $< 5 \mu\text{m}$ is about 28% less than the 'single crystal' starting material.

Our results are in reasonably good agreement with theory (Fig. 2) and confirm that in fine grained heavily doped silicon germanium alloys phonon scattering at grain boundaries results in a significant reduction in the lattice thermal conductivity. The implication is that thermoelectric generators which employ thermoelements fabricated from fine grained semiconductor alloys should exhibit substantial improvement in performance over those based on 'single crystal' or large grained materials.

We thank the British Gas Corporation for partial support of the project, Professor D. J. Harris for use of departmental facilities, and Dr R. D. Nasby for supplies of alloy.

Received 20 November 1980; accepted 27 February 1981.

1. Bhandari, C. M. & Rowe, D. M. *Contemp. Phys.* **21**, 219–242 (1980).
2. Rowe, D. M. *J. appl. Phys.* **7**, 1843–1846 (1974).
3. Goldsmid, H. J. & Penn, A. W. *Phys. Lett.* **27A**, 523 (1968).
4. Suzuki, K. & Maki, M. *J. appl. Phys. Jap.* **16**, 667 (1977).
5. Nasby, R. D. & Burgess, E. L. *Proc. 7th IECEC*, New York, 130 (1972).
6. Meddins, H. R. & Parrott, J. E. *J. Phys. C. Solid St. Phys.* **9**, 1263 (1976).
7. Bhandari, C. M. & Rowe, D. M. *J. Phys. C. Solid St. Phys.* **11**, 1787–1794 (1978).
8. Mullin, J. W. *Chem. Ind.* 1435–1436 (1971).
9. Bunce, R. W. & Rowe, D. M. *J. Phys. D. Appl. Phys.* **10**, 941–947 (1977).
10. Rowe, D. M. & Savvides, N. *J. Phys. D. Appl. Phys.* **12**, 1613–1619 (1979).
11. Taylor, R. J. *J. Phys. E. Scient. Instrum.* **13**, 1193–1199 (1980).
12. Dismukes, J. P. *et al. J. appl. Phys.* **35**, 2899–2907 (1964).
13. Rowe, D. M. *J. Phys. D. Appl. Phys.* **8**, 1042–1103 (1975).

Estimate of annual variation of Atlantic Ocean heat transport

Peter J. Lamb

Cooperative Institute for Marine and Atmospheric Studies, The University of Miami/NOAA, Miami, Florida 33149, USA and * Climatology Section, Illinois State Water Survey, Champaign, Illinois 61820, USA

Despite the importance of meridional oceanic heat transport for the understanding of the climate system^{1,2}, most existing hemispheric and ocean basin estimates are necessarily indirect. Annual averages have been obtained for individual basins from sea–air heat exchange results^{3–7}, and for each hemisphere from measurements of net radiation at the top of the atmosphere and calculations of atmospheric energy transport^{8,9}. The latter approach, when used in conjunction with atmospheric and oceanic energy storage information, has provided monthly mean estimates of the meridional heat transport by the total Northern Hemisphere water body, but not its separate oceans¹⁰. I now present an estimate of the annual variation of meridional heat transport in the Atlantic Ocean, based on large sets of sea–air interaction and subsurface temperature data, which suggests that, with the possible exception of November–December, the heat flux is northward.

Bimonthly averages of the 'vertically and zonally integrated net meridional heat transport' (VZINMHT) within the North and Tropical Atlantic Ocean (70° N–20° S, including the Gulf of Mexico and Caribbean Sea) were obtained for fixed latitude circle arcs spanning the entire basin (Table 1). The fundamental equation used was that for the heat budget of a local oceanic column,

$$Q_v = Q_{\text{stc}} - Q_t \quad (1)$$

where Q_v is the rate of lateral heat import/export by oceanic motions, Q_{stc} is the rate of net surface heat gain, and Q_t is the column's heat storage change rate. We computed average bimonthly integrals of these terms, with dimensions of Watts, for 10° (extratropics) and 5° (tropics) latitude zones extending across the basin. Those for Q_{stc} were obtained from existing sea–air heat exchange results. The Q_t integrals were estimated using a set of subsurface temperature data acquired and processed for this purpose. These results permitted Q_v to be obtained from equation (1). The zonal Q_v integrals were then converted into VZINMHT estimates (dimensions of Watts) for the zones' bounding latitude circle arcs, by successively integrating southwards from assumed near-zero 70° N VZINMHT boundary conditions computed from ref. 11 (Table 1). The uncertainty in the VZINMHT results given by this procedure naturally increases away from 70° N, and becomes large in the tropics. This is shown in Table 1. Such uncertainty arises from the use of empirical equations to estimate Q_{stc} , and from possible temporal and spatial sampling biases and instrumental inaccuracies in the data sets described below.

* Present address: Climatology Section, Illinois State Water Survey, Urbana, Illinois 61801, USA.

Table 1 Latitudinal variation of bimonthly average vertically and zonally integrated net meridional heat transport (VZINMHT) within the North and Tropical Atlantic Ocean, including the Gulf of Mexico and Caribbean Sea (10^{13} W)

Lat.	January–February		March–April		May–June		July–August		September–October		November–December		Bimonthly uncertainty	Average	Annual uncertainty
	300 m	500 m	300 m	500 m	300 m	500 m	300 m	500 m	300 m	500 m	300 m	500 m			
70° N	+07	+07	+07	+07	+07	+07	+15	+15	+15	+15	+09	+09	5	+10	5
60°	+85	+109	+18	−02	+14	+22	+36	+54	+24	+27	−06	−39	14	+28	8
50°	+126	+171	+57	+32	+29	+40	+59	+78	+50	+73	−15	−86	26	+52	12
40°	+79	+83	+83	+58	+82	+117	+128	+162	+90	+139	−31	−129	42	+73	17
30°	+101	+124	+107	+58	+126	+164	+148	+174	+151	+216	−03	−106	63	+107	24
25°	+101	+136	+126	+81	+113	+147	+169	+185	+157	+230	−05	−117	76	+113	29
20°	+85	+139	+137	+86	+118	+170	+187	+195	+152	+227	−13	−148	90	+114	33
15°	+83	+125	+149	+101	+121	+172	+189	+194	+165	+256	−33	−174	102	+115	37
10°	+90	+128	+197	+157	+75	+118	+180	+191	+162	+246	−45	−181	114	+113	41
5° N	+86	+120	+226	+193	+27	+60	+180	+201	+163	+241	−37	−168	123	+111	44
0°	+63	+79	+196	+164	+18	+61	+180	+203	+211	+295	−77	−209	133	+102	48
5° S	+50	+65	+137	+100	−04	+47	+170	+203	+253	+322	−107	−238	143	+86	51
10°	+59	+77	+138	+108	−13	+35	+156	+194	+257	+319	−121	−256	152	+82	53
15°	+49	+76	+142	+121	−14	+23	+150	+187	+257	+331	−106	−259	161	+83	57
20°	+57	+77	+111	+108	+01	+49	+152	+182	+260	+322	−115	−271	170	+81	60

Northward transports are positive; those for 70° N are boundary conditions computed from Aagaard and Greisman¹¹. Vertical integrations to 300 m and 500 m depths are given separately. The domain treated has the following noncontinental lateral boundaries: Baffin Island east coast, 64° W in Hudson Strait, 66° W in Gulf of St Lawrence, 6° W in Strait of Gibraltar, and 8° E in Skagerrak Strait. The bimonthly uncertainties (10^{13} W) are based on possible errors of 15 W m^{-2} in both Q_{sfc} and Q_e (refs 6, 10) and 5×10^{13} W in the 70° N boundary transports. The annual average VZINMHTs listed are computed from annual average Q_{sfc} s; they are very similar to the annual means of the bimonthly VZINMHTs cited, due to the annual average Q_e being extremely close to zero. The annual uncertainties (10^{13} W) are based on possible errors of 10 W m^{-2} in the annual average Q_{sfc} and 5×10^{13} W in the 70° N boundary transport.

The bimonthly average zonal Q_e integrals were estimated from 233,957 individual temperature soundings, using a 5° square spatial resolution. January–February values were obtained from the difference between the mean heat storage for December–January and February–March, and those for other bimonthly periods were calculated similarly. The temperature soundings used were drawn from the 1967–76 files of a National Oceanographic Data Center (Washington DC) global set which has been fully described by Levitus and Oort¹². Their spatial density is greatest between 30° and 40° N, from which it decreases both northwards and southwards (Table 2). Because of uncertainty about the representativeness of the sub-250 m annual temperature variations¹⁰ (only 54% and 20% of the soundings reach 300 m and 500 m, respectively), Q_e was computed for both the upper 300 m and upper 500 m. Table 1 contains VZINMHT estimates derived from each set of Q_e results.

The bimonthly average zonal Q_{sfc} integrals were calculated from the monthly mean sea–air heat exchange results of Bunker (70°–30° N) and Hastenrath and Lamb¹³ (30° N–20° S). Both these studies obtained long-term averages of Q_{sfc} from the surface heat budget equation

$$Q_{\text{sfc}} = \text{SW}\uparrow - \text{LW}\uparrow - Q_s - Q_e \quad (2)$$

The right-hand terms, which denote the net shortwave ($\text{SW}\uparrow$) and net longwave ($\text{LW}\uparrow$) radiation and the turbulent fluxes of sensible (Q_s) and latent (Q_e) heat, were estimated using conventional empirical equations and several million surface marine meteorological observations obtained from the National Climate Center (Asheville, NC). The data bases, methods, and uncertainties of each investigation have been fully discussed previously^{13–16}. Bunker's results, which are for Marsden (that is 10°) square subareas of varying size¹⁴, were derived from turbulent fluxes computed using the individual observations and variable drag coefficients necessary for the extratropics. Hastenrath and Lamb's results were preferred to Bunker's in the tropics because of the apparent superiority of their $\text{SW}\uparrow$ input¹⁶, and the fact that their 1° square spatial resolution permitted the use of 5° latitude zones in this energy source region for the climate system. Bunker's format required 10° zones.

The most important annual average VZINMHT result in Table 1 is that this flux is directed northwards throughout the study region. It increases from 81×10^{13} W at 20° S to a broad 5°–25° N maximum of 111 – 115×10^{13} W, due largely to substantial enhancement between 5° S and the Equator. There is a pronounced decrease poleward of 30° N, especially between 30° and 40° N. The general pattern of these results is very similar

to that recently obtained by Hastenrath⁶, whereas earlier work found the mean annual VZINMHT to have a northwards maximum near 30° N^{3,5}, weak northwards^{3,4}, or near zero⁵ cross equatorial values, and a southward direction throughout almost all of 0°–20° S (ref. 5). From 40° N southwards, however, Hastenrath's⁶ northwards VZINMHTs are 40–60% larger than the present ones. This difference primarily results from Hastenrath's 40°–50° N Q_{sfc} (taken from Budyko¹⁷) being more negative than the Bunker value used here. Budyko¹⁷ failed to detect the warming of the Labrador current between Nova Scotia and Cape Hatteras Identified by Bunker and Worthington¹⁵. The present annual average VZINMHT for 25° N ($+113 \times 10^{13}$ W) is in good agreement with Bryden and Hall's¹⁸ recent direct computation from oceanographic measurements ($+110 \pm 29 \times 10^{13}$ W).

Many of the bimonthly VZINMHT estimates in Table 1 possess considerable uncertainty, which requires that they be treated with some caution. Their most important result is the suggestion that northward transports may prevail throughout almost all of the study region during January–October. The northward flux out of the Northern Tropics seems to be strongest during July–October, when (particularly in September–October) the northwards VZINMHT may also be largest at 20° S and the Equator. This pattern is strongly enhanced for September–October when the storage computations are extended to 500 m. The northwards VZINMHT evidently undergoes only a relatively small reduction in the Northern Tropics during July–October, whereas much more pronounced decreases are indicated for the 40°–50° N (July–August) and 30°–50° N (September–October) zones. May–June, in contrast, seems to be characterized by a northward VZINMHT that is very weak at 20° S and the Equator, increases strongly between 0° and 15° N, approaches that of July–August at 30° N, then decreases uniformly to 50° N.

The most significant March–April result is the pronounced increase in the northward VZINMHT suggested to occur between 5° S and the Equator (Table 1). This converts a moderate northward flux at 20° S into an equatorial one comparable with that of July–October, and seems primarily responsible for the marked 5° S–0° increase in the annual average VZINMHT (Table 1). During this bimonthly period, the 0°–5° S zone experiences both a moderate positive Q_{sfc} and pronounced cooling in the upper 300 m (not shown). A progressive decrease of the March–April transport occurs north of 5° N. Table 1 suggests January–February is characterized by a northward VZINMHT that is quite small at 20° S, undergoes a moderate increase to 30° N, and may be stronger than other bimonthly periods at 50° and 60° N. The latter feature, along

Table 2 Number of 1967–76 subsurface temperature soundings available for 10° latitude zones of the Atlantic (including the Gulf of Mexico and Caribbean Sea), and the area of ocean surface (10^{12} m²) in each zone

Latitude zone	Ocean surface area	No. of soundings
60°–70° N	2.94	10,184
50°–60°	4.19	21,913
40°–50°	5.31	45,869
30°–40°	6.95	79,119
20°–30°	8.80	41,233
10°–20°	8.00	21,223
0°–10° N	6.64	6,763
0°–10° S	6.16	4,943
10°–20° S	6.12	2,710

The noncontinental lateral boundaries of the domain treated are specified in Table 1.

with the VZINMHT relative minimum at 40° N, arise from the North Atlantic having a maximum negative Q_{sc} between 30° and 40° N at a time when its cooling is strongly concentrated in the 40°–50° N zone (not shown). However, the comparative lack of winter subsurface temperature data for this region (only 44% of summer amount between 40° and 60° N) may render the foregoing results suspect.

The remaining significant feature of Table 1 is that its November–December results are in marked contrast to those for the rest of the year. The VZINMHT is suggested to be directed southwards throughout much or even all of the study region during November–December, and to show a general increase in this direction to a maximum at 10°–20° S. This is particularly indicated by the estimates derived from 500 m Q_t integrals, which have substantial magnitude south of 40° N. Much weaker, and extremely uncertain, southward VZINMHTs were yielded by the 300 m Q_t calculations, especially north of the Equator (Table 1). These November–December results arise from the oceanic cooling, especially when estimated to 500 m, exceeding the negative Q_{sc} north of 10° N (except for 30°–40° N), and from positive Q_{sc} values coinciding with lesser oceanic warming or oceanic cooling south of 10° N (not shown).

Although the November–December results are rather surprising and therefore of potential interest, they should be regarded as highly tentative. In particular, there does not seem to be a fully acceptable rationale for the southward VZINMHT currently available. Furthermore, the higher latitude Q_t and Q_{sc} estimates, on which all VZINMHTs depend, have a poorer data base for the cooler months than the rest of the year. Some relevant statistics were given above. If the November–December VZINMHT is in reality directed northwards rather than southwards, however, the present results probably imply that such northward transport is weaker than in the other bimonthly periods. The November–December results in Table 1 may therefore serve as a challenge to theoreticians and further stimulus for the planned programme to monitor the Atlantic VZINMHT near 25° N (refs 19, 20) directly.

I thank Eric B. Kraus for enabling me to undertake this research, and Stanley A. Changnon Jr for assistance. It was supported by NSF grant ATM 77 28126 to CIMAS, the State of Illinois, and the University of Adelaide Study Leave Programme. The late Andrew Bunker supplied sea–air heat exchange results, and Henry Poor gave computer support. I thank E. Kraus, C. Rooth, J. Merle, S. Hastenrath and W. Wendland for helpful comments.

Received 14 October 1980; accepted 18 February 1981.

- Interdepartmental Committee for Atmospheric Sciences A United States Climate Program Plan, 41–44 (Federal Coordinating Council for Science, Engineering, and Technology, Washington DC, 1977).
- US Committee for GARP Elements of the Research Strategy for the United States Climate Program 35–39 (National Academy of Sciences, Washington DC, 1978).
- Sverdrup, H. U. *Handbuch der Physik* Vol. 48, 608–670 (Akademie, Berlin, 1957).
- Bryan, K. J. *geophys. Res.* **67**, 3403–3414 (1962).
- Emig, M. J. *geophys. Res.* **72**, 2519–2529 (1967).
- Hastenrath, S. J. *phys. Oceanogr.* **10**, 159–170 (1980).
- Hastenrath, S. & Lamb, P. J. *J. phys. Oceanogr.* **10**, 694–708 (1980).
- Vonder Haar, T. H. & Oort, A. H. *J. phys. Oceanogr.* **3**, 169–172 (1973).
- Trenberth, K. E. *Dynamics atmos. Oceans*, **4**, 57–64 (1979).
- Oort, A. H. & Vonder Haar, T. H. *J. phys. Oceanogr.* **6**, 781–800 (1976).

- Aagaard, K. & Greisman, P. J. *geophys. Res.* **80**, 3821–3827 (1975).
- Levitus, S. & Oort, A. H. *Bull. Am. met. Soc.* **58**, 1270–1284 (1977).
- Hastenrath, S. & Lamb, P. J. *Heat Budget Atlas of the Tropical Atlantic and Eastern Pacific Oceans* (The University of Wisconsin Press, Madison, 1978).
- Bunker, A. F. *Mon. Weath. Rev.* **104**, 1122–1140 (1976).
- Bunker, A. F. & Worthington, L. V. *Bull. Am. met. Soc.* **57**, 670–678 (1976).
- Bunker, A. F. *Bull. Am. met. Soc.* **60**, 131–132 (1979).
- Budyko, M. I. *Atlas of the Heat Balance of the Earth* (Kartfabrika Gosgeoltekhizdat, Leningrad, 1963).
- Bryden, H. L. & Hall, M. M. *Science* **207**, 884–886 (1980).
- National Oceanic and Atmospheric Administration An Ocean Climate Research Plan, 25–28 (US Department of Commerce, Washington DC, 1979).
- National Oceanic and Atmospheric Administration Ocean Heat Transport and Storage, 5–22 (US Department of Commerce, Washington DC, 1979).

Dissolved state of chromium in seawater

Eiichiro Nakayama, Hiroyuki Tokoro,
Tooru Kuwamoto & Taitiro Fujinaga

Department of Chemistry, Faculty of Science, Kyoto University,
Kyoto 606, Japan

The dissolved state of chromium in seawater has long been studied and in particular which species, trivalent or hexavalent, is predominant. Research results are not, however, consistent, although the method used has usually been the same—using ferric hydroxide as coprecipitation carrier (Table 1^{1–7}). We have studied the coprecipitation behaviour of chromium with ferric hydroxide and other metal hydroxides in the presence of naturally existing inorganic and organic ions to attain an accurate analytical method for determining chromium content, as well as to estimate its behaviour in natural waters^{8,9}. We report here that neither Cr(VI) nor organic species coprecipitate with ferric hydroxide in seawater; Cr(VI) can be quantitatively captured by the coprecipitation with bismuth hydroxide without any pretreatment (such as reduction); and the role of manganese oxide should be considered in addition to the amount of dissolved oxygen to understand the redox system of chromium in natural waters. The inconsistency of the past research may therefore be partly because the presence of organic species was not taken into account and was analysed as either Cr(VI) or Cr(III) or overlooked entirely.

We have developed a method for the fractional quantitative determination of chromium in seawater. A 1.2-l sample of seawater was filtered through a 0.4- μ m nucleopore filter immediately after which the sample was divided into three. To the first and second portions were added respectively ferric hydroxide and bismuth hydroxide precipitates, which had been prepared separately and thoroughly washed with distilled water. In seawater solution with a pH ~8 inorganic Cr(III) only coprecipitates quantitatively with ferric hydroxide, inorganic Cr(III) and Cr(VI) coprecipitate with bismuth hydroxide while the organic species hardly precipitate at all with either hydroxide. The third portion was acidified with hydrochloric acid and heated with ammonium persulphate to decompose any organic matter. The pH of this portion was then adjusted back to 8 with aqueous ammonia and bismuth hydroxide was added. After these portions had been allowed to stand for 12 h, the precipitates were filtered and dissolved in hydrochloric acid. The iron and bismuth was removed by the solvent extraction method using a *p*-xylene solution of 5% tri-isooctylamine, the aqueous phase evaporated to dryness, and the residue dissolved in a small amount of diluted nitric acid. The resulting solution was used for flameless atomic absorption analysis. The total amount of inorganic Cr(III) present was determined from the first portion; the total inorganic Cr(III) and Cr(VI) from the second portion; and the total amount of chromium, including the organic species from the third portion. By performing the experiment with a ⁵¹Cr tracer, it was found that the recovery of chromium was >95% in the coprecipitation procedure and >99% in the other procedures. The procedure was repeated five times with each

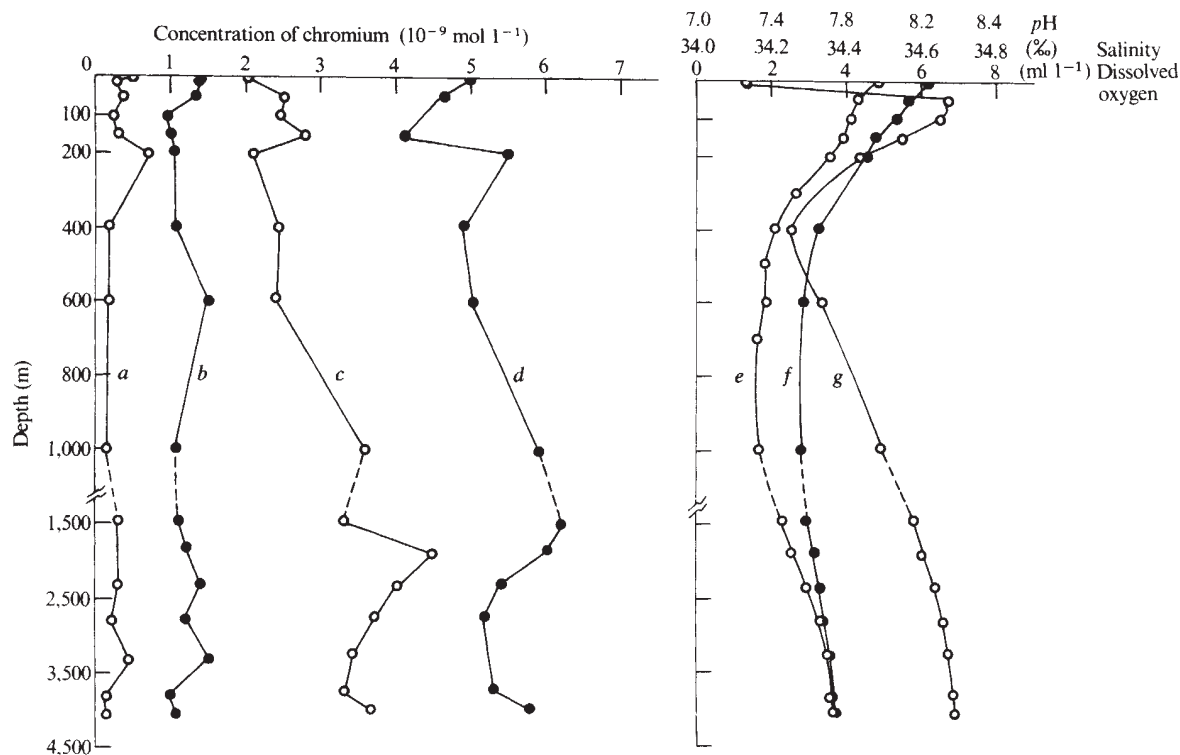


Fig. 1 Vertical distribution of chromium species and some standard components in the Pacific Ocean ($32^{\circ}19.3'N$, $137^{\circ}33.5'E$; depth 4,079 m). *a*, Particulate chromium; *b*, dissolved inorganic Cr(III); *c*, dissolved Cr(VI); *d*, organic chromium species; *e*, dissolved oxygen; *f*, pH; *g*, salinity.

sample and the standard deviations for the first, second and the third portions were 5.8, 2.4 and 5.1% respectively. The amount of particulate chromium present was determined by analysing the filter paper used for filtrating the sample.

Although there was no particular quality control to minimize chromium contamination, seawater sampling was usually carried out in the direction of the wind and plastic Niskin-type bottles were used to avoid metal contamination in our research vessel.

This method was used to analyse the samples of seawater collected in the adjacent seas of Japan in the summer of 1979. Figures 1 and 2 show typical results from six stations in the Pacific Ocean and four stations in the Japan Sea, respectively. The total average concentration of dissolved chromium is $9.4 \times 10^{-9} M$ ($0.48 \mu g l^{-1}$) in the Pacific Ocean and $8.1 \times 10^{-9} M$ ($0.42 \mu g l^{-1}$) in the Japan Sea: consisting of 10–20% of inorganic Cr(III), 25–40% of Cr(VI) and 45–65% of organic species. The average concentration of particulate chromium is

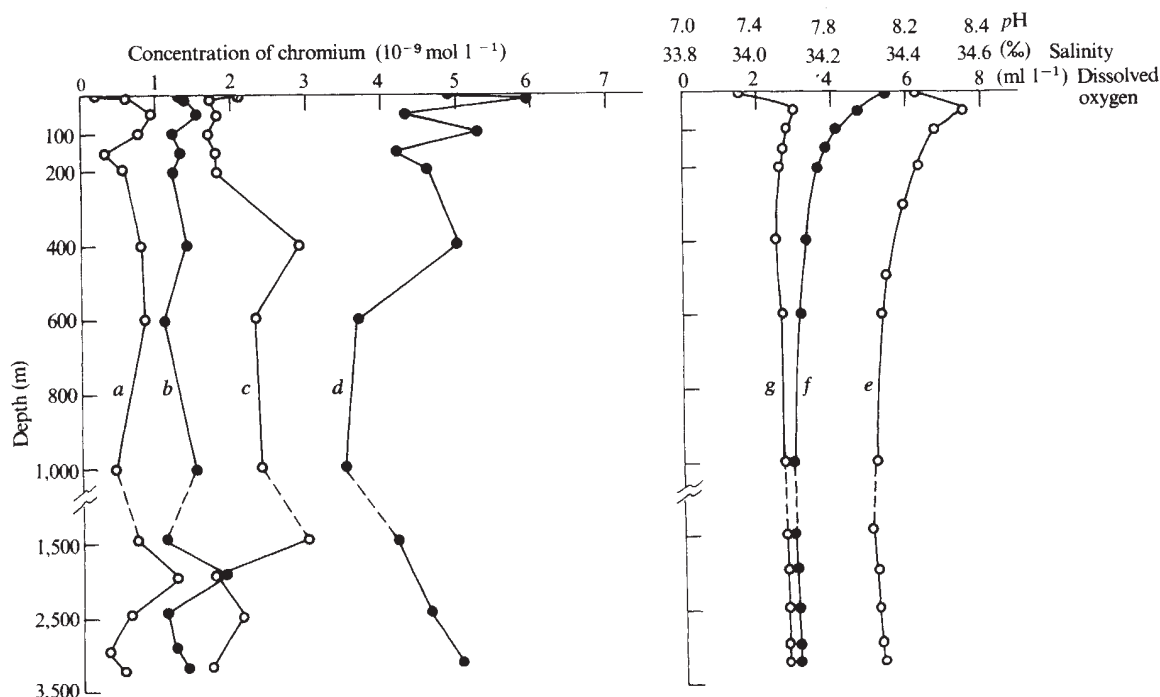


Fig. 2 Vertical distribution of chromium species and some standard components in the Japan Sea ($44^{\circ}11.9'N$, $138^{\circ}56.4'E$; depth 3,447 m). *a*, Particulate chromium; *b*, dissolved inorganic Cr(III); *c*, dissolved Cr(VI); *d*, organic chromium species; *e*, dissolved oxygen; *f*, pH; *g*, salinity.

Table 1 Literature data on the chromium contents in seawater

Ref.	Locations	Concentration of chromium ($\mu\text{g l}^{-1}$)	Ratio of Cr(VI)/Cr(III)
1	British coastal	Cr(III) 0.46 Cr(VI) 0	Cr(III) only
2, 3	Ligurian Sea	Cr(III) 0.02–0.25 Cr(VI) 0.05–0.36	Cr(VI) > Cr(III)
4	Pacific Ocean	Cr(III) 0.13–0.90 Cr(VI) 0.07–0.27	Cr(VI) < Cr(III)
5	Equatorial region of Pacific Ocean	Cr(III) 0.24–0.52 Cr(VI) 0.03–0.11	Cr(VI) < Cr(III)
6	Northern north Pacific Ocean	Total Cr 0.16–0.95 Cr(VI) 5–70%	Cr(VI) < Cr(III)
7	North-east region of Pacific Ocean	Cr(III) 0.005 Cr(VI) 0.15	Cr(VI) > Cr(III)

3.6×10^{-10} M ($0.07 \mu\text{g l}^{-1}$) in the Pacific Ocean and 4.2×10^{-10} M ($0.08 \mu\text{g l}^{-1}$) in the Japan Sea. Figures 1 and 2 show that the vertical distribution of Cr(VI) differs between the two Seas. In the Pacific Ocean there is a remarkable increase at a depth of 1,000 m and below while in the Japan Sea there is a decrease. The ratio of Cr(VI) to inorganic Cr(III) averages 2.7 in the Pacific Ocean while in the Japan Sea it averages 1.8. This difference cannot be explained by thermodynamical equilibrium and application of the oceanographical data, such as pH and dissolved oxygen¹⁰, because the two seas are completely oxygenated. Hence we propose a model for the circulation of chromium in seawater as shown in Fig. 3. This model was supported by an experiment based on the geochemical fact that in manganese nodules the chromium content is extremely low compared with that in the marine sediments composed of clay minerals¹¹. Thus, when the seawater solution containing 10^{-5} M of Cr(III) was bubbled with air for >300 h no hexavalent chromium was detected. However, when about 50 mg l^{-1} of powdered manganese oxide (such as synthesized manganite and manganese nodules) were added, 10% of Cr(III) was oxidized within 100 h. Organic complexes of Cr(III) (such as chromic citrate) were not oxidized within 100 h even if excess manganese oxide was present. From this model, the behaviour of chromium in seawater can be explained as follows. When inorganic Cr(III) is adsorbed by suspended particles it is oxidized by the catalytic action of manganese oxide during sedimentation or in the marine sediments. Direct oxidation with dissolved oxygen only is not appreciable. Cr(VI) is reduced by being taken into organisms or by reductive organic matter. When organic species form, oxidation ceases.

Thus we interpret the difference of distribution of Cr(VI) between the two Seas as being due to the fact that the marine sediments contain abundant manganese oxide in the Pacific Ocean because they are oxygenated down to a considerable depth, whereas the sediments in the Japan Sea, except for a thin oxygenated layer at the surface reduce manganese oxide to Mn(II) (ref. 12). Accordingly, Cr(VI) is supplied to the water

column in the Pacific Ocean, but not to the Japan Sea. In addition, the Cr(VI) supplied from suspended particles does not seriously effect its vertical distribution because their chromium content is very low.

We thank members of the staff of the Ocean Research Institute, University of Tokyo for their cooperation in the collection of samples.

Received 25 September 1980; accepted 24 February 1981.

1. Chuecas, L. & Riley, J. P. *Analyt. chim. Acta* **35**, 240–246 (1966).
2. Fukai, R. *Nature* **213**, 901 (1967).
3. Fukai, R. & Vas, D. J. *oceanogr. Soc. Jap.* **23**, 298–302 (1967).
4. Kuwamoto, T. & Murai, S. *Preliminary Rep. Hakuho-maru Cruise KH-68-4*, 72 (Ocean Research Institute, University of Tokyo, 1970).
5. Grimaud, D. & Michard, G. *Mar. Chem.* **2**, 229–237 (1974).
6. Yamamoto, T., Kadowaki, S. & Carpenter, J. H. *Geochem. J.* **8**, 123–133 (1974).
7. Cranston, R. E. & Murray, J. M. *Analyt. chim. Acta* **99**, 275–282 (1978).
8. Fujinaga, T., Kuwamoto, T., Murai, S., Kihara, S. & Nakayama, E. *Nippon Kagaku Zasshi* **92**, 339–344 (1971).
9. Fujinaga, T., Kuwamoto, T., Nakayama, E. & Tsurubo, S. *Abstr. Papers ACS/CSJ Chem. Congr. Envir. Chem. No. 100* (Honolulu, Hawaii, 1979).
10. Elderfield, H. *Earth planet. Sci. Lett.* **9**, 10–16 (1970).
11. Turekian, K. K. *Chemical Oceanography* Vol. 2, 81–125 (Academic, New York, 1955).
12. Masuzawa, T. *Abstr. oceanogr. Soc. Jap.*, 148–149 (1978).

Effect of irregular fluctuations in Antarctic precipitation on global sea level

J. Oerlemans

Institute of Meteorology and Oceanography, University of Utrecht, Postbus 80.005, 3508 TA Utrecht, The Netherlands

One of the reasons for the continuing interest in the global sea level is that secular variations may be caused by climatic changes. Such a change could, for example, be an atmospheric warming due to CO_2 accumulation. Changes in the amount of ice in the major ice sheets will be reflected in secular variations of sea level; it has, for example, been suggested that ice-shelf thinning may change the drainage of parts of the Antarctic Ice Sheet¹. Attempts to monitor climatic change by measuring global sea level will, however, be complicated by random fluctuations of the ice volume in the major ice sheets, themselves the consequence of random variations in the ice accumulation rate. Precipitation rates are highly variable, and this also applies to Antarctica², which stores most of the continental ice mass. By means of a simple model for ice flow in the Antarctic, together with proxy data on precipitation variability derived from ice cores, I show that long-term sea-level variations with a standard deviation of roughly 5 cm are to be expected on this account. This 'climatic noise' is comparable in magnitude with many of the secular effects now being sought.

Any changes in the volume of the Antarctic Ice sheet are damped by a mechanism involving several drainage systems³. In these systems ice is discharged to the ocean by horizontal pressure gradients resulting from differences in elevation of the ice surface. If the ice thickness of a drainage system is increased (without changing the mass balance—the yearly gain of ice mass at the surface), the mean pressure gradient from the interior of the drainage system to the ocean will become larger. The subsequent increase in ice-mass discharge will try to restore the original ice thickness. In the following, we will measure the strength of this damping by the relaxation time, t_R , which indicates how fast the system can return to equilibrium.

Once this time scale is known, the effect of random fluctuations in the mass balance on the volume of the Antarctic Ice Sheet can be estimated with a simple stochastic model. A framework for the application of stochastic models to climate problems has been presented by Hasselmann⁴. Here I will use a simple model, consisting of a rate equation for ice thickness with a linear damping term and an additive stochastic term to represent the fluctuations in the ice accumulation rate.

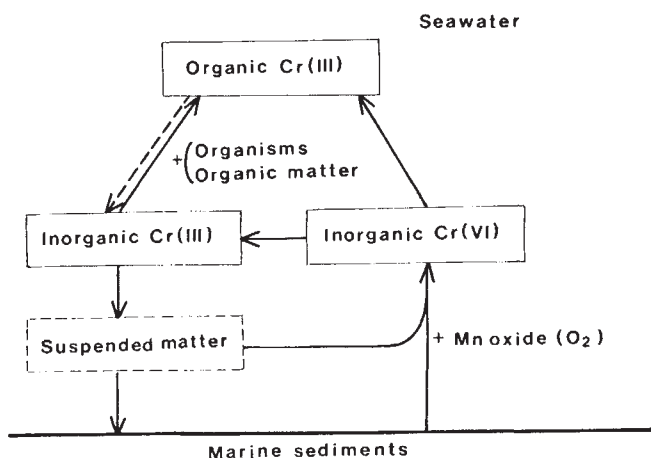


Fig. 3 A model for the circulation of chromium in seawater.

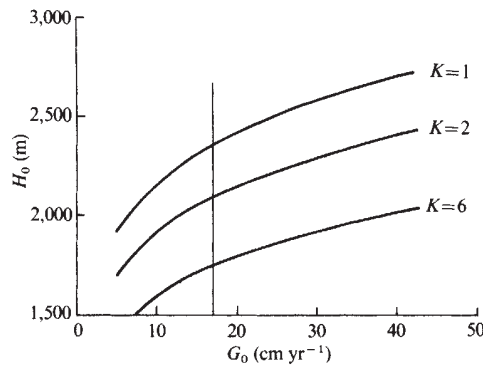


Fig. 1 Steady-state ice thickness H_0 , as a function of the mass balance G_0 , for three values of the flow parameter K ($\text{m}^{-3/2} \text{yr}^{-1}$). The vertical line indicates the present mass balance over Antarctica.

To obtain a value for t_R , we can turn to a schematic physical description of the dynamics of large ice streams. As proposed by Nye⁵, we use a flow law for the vertically integrated horizontal ice velocity (denoted by u), which should account for both internal deformation and (moderate) basal sliding. It reads

$$u = k\tau_b^m \quad (1)$$

where τ_b is the basal shear stress and k and m are constants. The basal shear stress is proportional to the ice thickness, H , times the slope of the ice surface. For simplicity, we assume that the bedrock is flat. After some manipulations, a one-dimensional ice-sheet model based on the conservation of ice mass, can be formulated as follows

$$\frac{\partial H}{\partial t} = \frac{\partial}{\partial x} \left[D \frac{\partial H}{\partial x} \right] + G(x) \quad (2)$$

where

$$D = kH^{m+1} \left| \frac{\partial H}{\partial x} \right|^{m-1}.$$

The x axis is oriented in the direction of the mean ice velocity in the drainage system, that is, from the interior of the Antarctic continent to the ocean. The mass balance is denoted by $G(x)$. I have used the model formulated by equation (2) in some other studies. (See ref. 6 for details.) Current values for m are within the 2–3.5 range. Here we adopt a value of 2.5. The flow parameter, k , will be chosen such that the present model produces the correct order of magnitude of ice thickness in a typical Antarctic drainage system.

To make the model more tractable, we will now make it zero-dimensional by inserting in equation (2) typical values of the quantities involved. Let the typical ice thickness be H_0 and the typical horizontal scale for ice thickness variations L_0 . To a first approximation, the dynamics of the drainage system can be described by

$$\frac{dH_0}{dt} = -KH_0^{2m+1}L_0^{-(m+1)} + G_0 \quad (3)$$

The constant, K , is not necessarily the same as k . The steady-state ice thickness can easily be found by setting the dH_0/dt to zero. The resulting relation between H_0 and G_0 is displayed in Fig. 1. L_0 has been set to 1,000 km. Curves are shown for $K = 1, 2$ and $6 \text{ m}^{-3/2} \text{yr}^{-1}$, corresponding to $H_0 = 2,311, 2,055$ and $1,712 \text{ m}$ respectively, for $G_0 = 0.17 \text{ m}$ of ice per yr (which is about the average mass balance of the Antarctic Ice Sheet).

To fit the description of ice-flow dynamics given above into a simple linear stochastic model, equation (3) has to be linearized around some reference state. For this purpose, it is natural to use the present state (although we cannot be sure that at present the Antarctic Ice Sheet is close to equilibrium), which will be

denoted by asterisks. We find

$$\frac{dH'_0}{dt} = -K(2m+1)H_0^{*2m}L_0^{-(m+1)}H'_0 + G'_0 \quad (4)$$

A prime denotes deviation from the reference state. According to equation (4), the relaxation time t_R equals $[K(2m+1)H_0^{*2m}L_0^{-(m+1)}]^{-1}$. Values of t_R corresponding to $H_0^* = 2,353, 2,097$ and $1,746 \text{ m}$ are 2,311, 2,055 and 1,712 yr, respectively. Apparently, t_R is not very sensitive to the particular reference state we choose.

At this point, we interpret $G'_0(t)$ as a stochastic process, representing random fluctuations in the mass balance. The quantity H'_0 is assumed to represent deviations of the mean thickness of the Antarctic Ice Sheet (note that this implies that the various drainage systems have the same dynamical characteristics, which seems a reasonable approximation for this analysis). To keep things as simple as possible, we assume that $G'_0(t)$ is a white-noise process, which implies that the process $H'_0(t)$ has a spectrum proportional to $1/(1+\omega^2 t_R^2)$ (red noise). The total variance of $H'_0(t)$ equals $t_R S_G$, where S_G is the (constant) spectral density of the white noise⁴. In practice, a reliable estimate of S_G is best obtained from mass balance measurements that are averaged over some time. If the averaging time is t , the following relation results:

$$\sigma_{H_0} = (t t_R)^{1/2} \sigma_G \quad (5)$$

Here, σ denotes standard deviation and the tilde indicates that the mass balance is averaged over t yr. Choosing $t_R = 2,100 \text{ yr}$ (intermediate curve in Fig. 1) and $t_R = 30 \text{ yr}$, we have $\sigma_{H_0} = 250 \sigma_G$.

At this point an estimate of σ_G is needed. Figure 2 shows some proxy data on mass balance variations that were readily available and permit an order-of-magnitude estimate of σ_G . The records are inferred from stratigraphical studies at three locations on the Greenland Ice Sheet⁷ and at the South Pole⁸. The data shown suggest a relative standard deviation for 30-yr mean values of roughly 7%. For the Antarctic Ice Sheet with a mean mass balance of 0.17 m ice per yr, this gives $\sigma_G = 0.012 \text{ m}$ ice yr^{-1} . It cannot be expected that precipitation anomalies are uniform over Antarctica. To correct for this we need to know into how many 'independent' regions Antarctica can be divided. At present, no data are available on the typical horizontal scale of accumulation anomalies over the Antarctic Ice Sheet. We therefore arbitrarily assume that five regions exist where

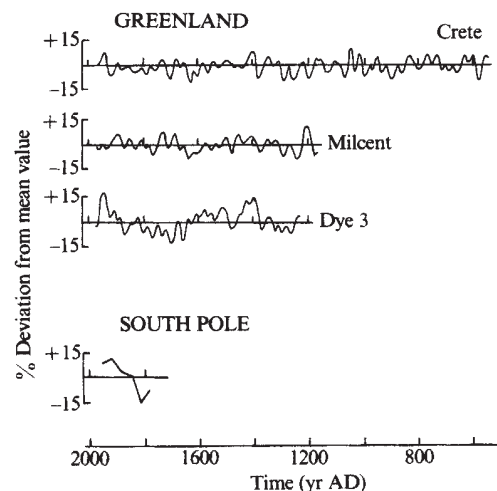


Fig. 2 Time series showing variations in ice accumulation rates, as inferred from stratigraphical studies on the ice of Greenland and Antarctica. The three upper curves are from ref. 7. Variability on time scales smaller than 30 yr has been removed from these curves. The lower curve shows 30-yr mean values of accumulation at the South Pole, computed from data given in ref. 8. Variability is expressed as relative deviation from the mean value.

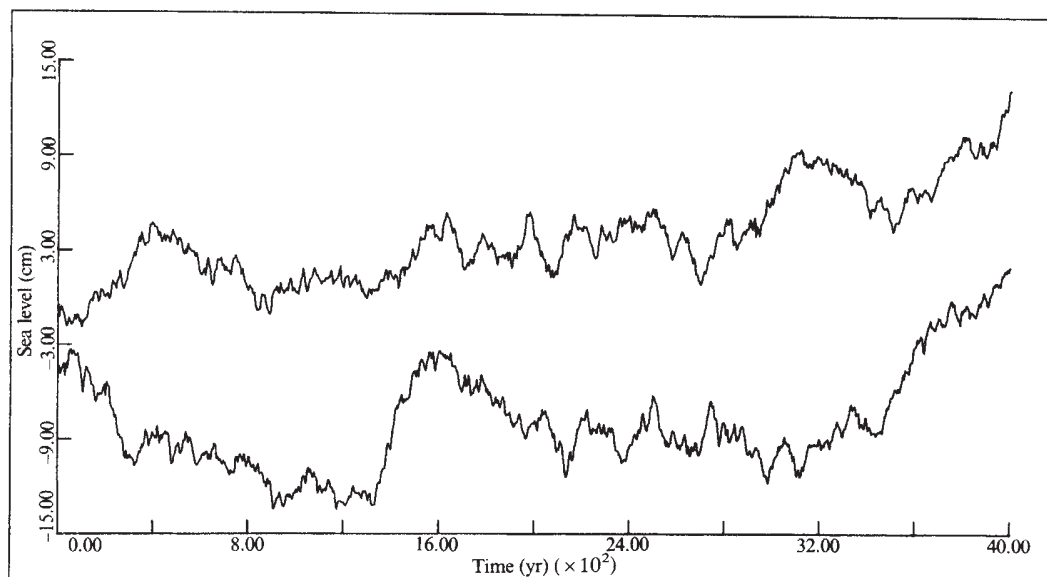


Fig. 3 Two sea-level records generated by the stochastic model, with a length of 4,000 yr. Values of the constants used: $t_R = 2,100$ yr, $\sigma_{\hat{G}} = 0.85$ cm ice per yr.

snowfall fluctuations are uncorrelated. The value of $\sigma_{\hat{G}}$ should then be divided by $\sqrt{5}$, which gives $\sigma_{\hat{G}} = 0.0054$ m ice yr^{-1} . The resulting value of σ_{H_0} is about 1.35 m. As the area of the Antarctic Ice Sheet is $\sim 4\%$ of that of the world ocean, the corresponding standard deviation of the sea level amounts to ~ 5 cm. If we would choose the spatial scale of precipitation much smaller, the value of σ_{H_0} would also be smaller. For example, if nine independent regions were present, σ_{H_0} would be about 1 m.

To show more clearly what happens, Fig. 3 provides two examples of sea-level records generated by equation (4). The process $G'_0(t)$ was simulated by a random number generator, such that $\sigma_{\hat{G}} = 0.008$ m ice yr^{-1} . The records, both spanning, 4,000 yr, show that the major part of the variability is contained in the longer time scales. This is to be expected, because the relaxation time of the ice sheet is of the order of a few thousand years.

The foregoing results clearly demonstrate that stochastic forcing of the Antarctic Ice Sheet by fluctuations in snow accumulation may contribute significantly to variations in global sea level. As revealed in Fig. 3, a sea-level trend of 4 cm in 100 yr can easily be due to the process analysed in this study. Even if local sea-level changes are considered, this is by no means negligible. Other mechanisms creating long-term sea-level variability (like changes in atmospheric wind and pressure patterns, shifts in the oceanic circulation regimes, eustatic effects) lead to changes with the same order of magnitude⁹.

It should be stressed that I have presented here an order-of-magnitude estimate. In the foregoing analysis several uncertainties exist. First, the estimate of $\sigma_{\hat{G}}$ is very crude. A critical survey of all data on accumulation over the Antarctic continent is needed to derive a typical horizontal scale for accumulation anomalies. Another point concerns the very schematic treatment of the dynamics of the Antarctic Ice Sheet. The geometry of Antarctica is rather complicated (large variations in bedrock elevation), so modelling of the ice flow by a zero-dimensional model can only give a crude approximation to reality. The assumption of constant L_0 is a poor approximation for regions where the Antarctic continent is not bounded by deep ocean but by a continental shelf, which makes it possible for the ice sheet to expand horizontally.

In view of these points, perhaps the vertical axis in Fig. 3 should be doubled, or halved. In spite of this uncertainty, however, this study reveals that a substantial part of global sea-level variability may be caused by random forcing of the Antarctic Ice Sheet.

Received 3 December 1980; accepted 18 February 1981.

1. Thomas, R. H., Sanderson, T. J. O. & Rose, K. E. *Nature* **277**, 355–358 (1979).
2. Schwerdtfeger, W. in *World Survey of Climatology* Vol. 14 (ed. Landsberg, H. E.) 253–355 (1970).
3. Giovinetto, M. B. in *Antarctic Snow and Ice Studies*, Antarctic Res. Ser. 2 (ed. Mellor, M.) 127–155 (American Geophysical Union, 1964).
4. Hasselmann, K. *Tellus* **28**, 473–485 (1976).
5. Nye, J. F. *J. Glaciol.* **3**, 493–507 (1959).
6. Oerlemans, J. *Tellus* **33**, 1–12 (1981).
7. Reeh, N. *et al. J. Glaciol.* **20**, 27–30 (1978).
8. Giovinetto, M. B. & Schwerdtfeger, W. *Arch. Meteorol. Geophys. Bioklimatol.*, Ser. A **15**, 228–250 (1966).
9. Lisitzin, E. *Sea-level Changes* (Elsevier, Amsterdam, 1974).

The Caledonian nappes of eastern North Greenland

John Malcolm Hurst* & William Stuart McKerrow†

* Grønlands Geologiske Undersøgelse, Øster Voldgade 10, 1350 Copenhagen K, Denmark

† Department of Geology and Mineralogy, Parks Road, Oxford OX1 3PR, UK

Reconnaissance work in Kronprins Christian Land^{1–7} indicated the presence of one main nappe of late Proterozoic sediments^{2,3} thrust over Lower Palaeozoic platform carbonates similar to Peary Land^{7,8}. Our 2-month helicopter-supported work in 1980 now indicates the presence of three thrust-bounded nappes. The Funderup Land Nappes, which occur in northern Kronprins Christian Land, contain a late Precambrian early Cambrian sequence in thrust contact above Silurian carbonates. The younger Vandredalen Nappe in thrust contact above the Funderup Land Nappes and Silurian carbonates occurs throughout Kronprins Christian Land and contains a thick sequence of late Proterozoic clastics and carbonates. In southern Kronprins Christian Land the Sæfaxi Elv Nappe only contains Ordovician–?Silurian carbonates and is in thrust contact below Vandredalen Nappe. The nappes were involved in westward-directed thrusting (≥ 150 km) and are probably a result of continental collision when the northern part of Iapetus Ocean closed in the late Silurian.

On the east coast of Kronprins Christian Land (Fig. 1), a narrow belt of gneisses crops out to the east of Proterozoic sandstones which have been referred previously to the Thule Group^{5,6,9}. Both are probably equivalent to the Proterozoic Independence Fjord Group west of Danmark Fjord¹⁰ (Fig. 1).

The Proterozoic sandstones are intruded by numerous dolerite dykes and sills. West of Danmark Fjord (Fig. 1) an intrusive granophyre in the Independence Fjord Group has yielded a whole-rock Rb-Sr isochron age of $1,230 \pm 25$ Myr (ref. 11) and radiometric dates (minimum age) of interbedded clays give a 1,380 Myr age¹². The Proterozoic sandstones of Kronprins Christian Land are probably of a similar age ($>1,000$ Myr).

The Proterozoic sandstones in Kronprins Christian Land are locally overlain by a conglomerate³ consisting of large (10–100 cm) angular to rounded boulders of Proterozoic sandstone and dolerite. This conglomerate is followed at Hjørnegletscher (E, Fig. 1) by a thin sequence of shallow marine clastics and a stromatolitic dolomite identical to the late Precambrian Fyns Sø Formation of the area west of Danmark Fjord; but at Marmorvigen (C, Fig. 1) the Fyns Sø Formation rests directly on the conglomerate.

The Proterozoic sandstones in Kronprins Christian Land are deformed by strong Caledonian movements (H. F. Jepsen and F. Kalsbeek, personal communication). In Ingolf Fjord the Proterozoic sandstones and the overlying late Precambrian shallow marine clastic sediments are involved in a series of broad anticlines and narrow inclined synclines, showing overriding movement to the west. South of Ingolf Fjord (Fig. 1) these folds are replaced by westward-directed thrusts of Proterozoic sandstone over Fyns Sø Formation. Although the Proterozoic sandstones are in part allochthonous in southern Kronprins Christian Land, they have not moved far, as the thrusts in the south pass laterally into folds in the north.

To the west of the Proterozoic sandstones three large thrust-bounded nappes (Fig. 1) (or groups of nappes) with distinctive

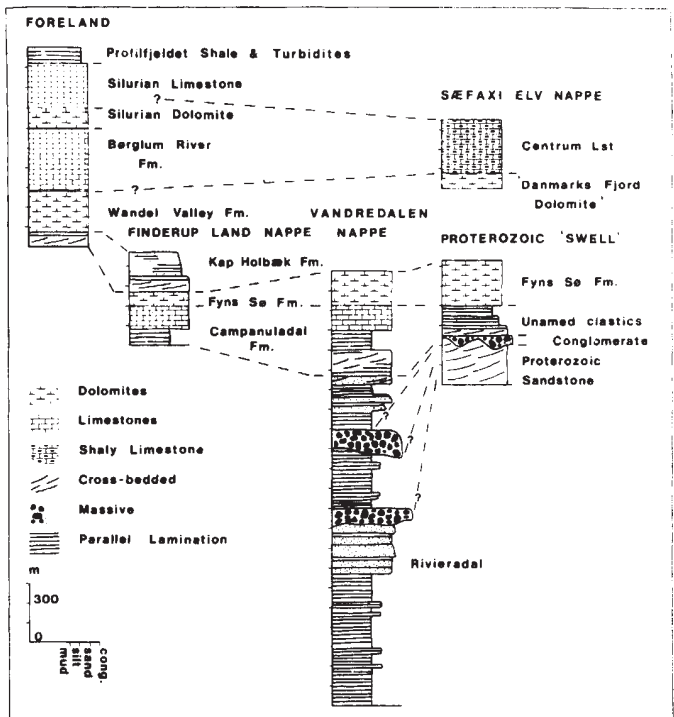


Fig. 2 Stratigraphy and facies distribution of the autochthonous (Foreland' and Proterozoic 'Swell') and allochthonous (FINDERUP Land Nappes, Vandredalen Nappe and Sæfaxi Elv Nappe) sediments.

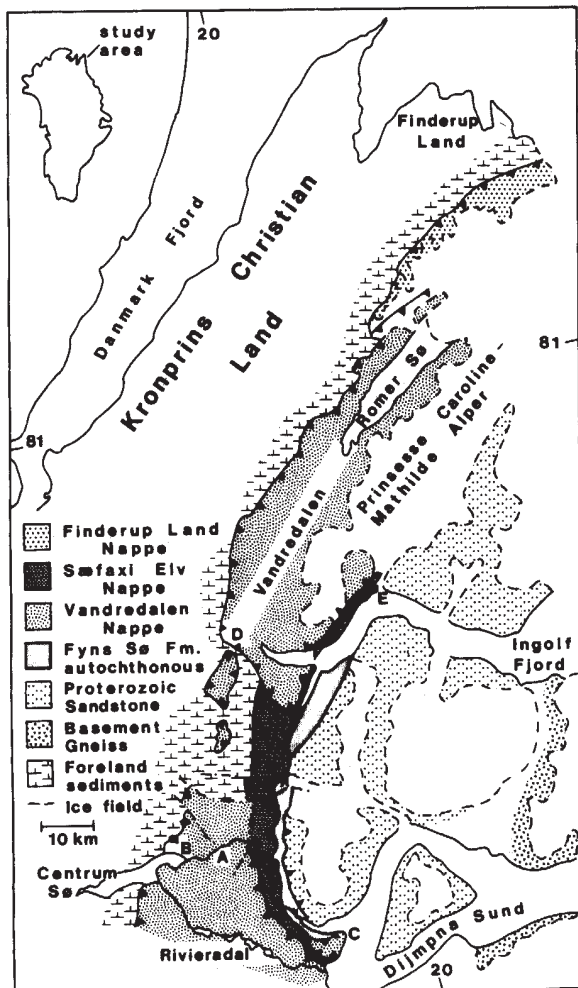


Fig. 1 Simplified geological map of Kronprins Christian Land, North Greenland.

stratigraphic sequences are present in Kronprins Christian Land: (1) The FINDERUP Land Nappes in the north (Fig. 1). These contain sequences of stromatolitic limestone and clastics of the Campanuladal Formation and deformed stromatolitic dolomite of the Fyns Sø Formation overlain by sandstones with *Skolithos* and bioturbated siltstones of the Cambrian Kap Holbæk Formation. (2) The Vandredalen Nappe of central and southern Kronprins Christian Land (Fig. 1). This consists of a thick sequence of turbidites (Rivieradal sandstones^{2,3}) overlain by sands, silts, variegated shales, silts and red limestones (in part Campanuladal Formation) and stromatolitic dolomites (Fyns Sø Formation). This stratigraphic sequence is well exposed on Drømmebjerg (B, Fig. 1), on Nøglefjeldet (D, Fig. 1) and other hills west of Vandredalen. Basal turbidites of the Rivieradal unit also occur on a higher nappe at Marmorvigen (C, Fig. 1) and similar silty turbidites may be on separate nappes in northern Kronprins Christian Land. (3) The Sæfaxi Elv Nappe in south-east Kronprins Christian Land (Fig. 1). This contains a sequence of dolomites with sandstone fissures which have been referred to the Danmarks Fjord dolomite^{2,3}, overlain by Ordovician to Silurian black limestones and limy shales of the Centrum Limestone unit.

In addition to these large thrust-bounded nappes, several complex imbricate zones are exposed along their eastern and western margins. To the south-east, near Marmorvigen and along the south-west side of the Sæfaxi Elv (A, Fig. 1), the Vandredalen Nappe rests on the Sæfaxi Elv Nappe and both are thrust over a sequence of Fyns Sø Formation and the conglomerate above the Proterozoic sandstones and dolerites.

The FINDERUP Land Nappes occur to the north of a large ice cap (Fig. 1) so that its relationship to the Vandredalen Nappe is not clearly exposed. However, the late Precambrian and Cambrian Fyns Sø Formation and Kap Holbæk Formation of the FINDERUP Land Nappes appear to underlie the Vandredalen Nappe, and this is borne out by the fact that the beds of this nappe are directly in thrust contact with the folded Silurian platform carbonates in FINDERUP Land.

To the west, the nappes are in thrust contact with the folded sequence of Ordovician and Silurian platform carbonates and

Silurian trough clastics⁷. In the vicinity of the thrusts these Palaeozoic rocks are locally overturned towards the west. If the strike of the fold axes is a direct response to the direction of thrusting, the movement of the Vandredalen Nappe was between 260° and 280°, while in the north the Finderup Land Nappes seem to have travelled towards 310°. To the west and north-west of the nappes, the Palaeozoic beds show asymmetrical folds with steep or overturned north-west or west facing limbs to the anticlines. This disturbed zone stretches for up to 40 km west of the present exposures of the nappes⁷ and, if it is due to deformation by the overriding nappes, it is possibly an indication that the nappes may have formerly extended to Danmark Fjord. Although some steeply inclined reversed faults occur in this region of disturbed Palaeozoic strata, none of the beds has been transported far; they are essentially autochthonous⁷. To the west of Danmark Fjord the same Palaeozoic platform strata are flat-lying and rest undisturbed on earlier Precambrian sediments.

The stratigraphic sequences (Fig. 2) of the large nappes and of the Proterozoic sandstones to the east can only be correlated by: (1) The presence of the stromatolitic dolomites of the Fyns Sø Formation which are known to be late Precambrian on the platform to the west. (2) The presence of *Skolithos* and bioturbated silts which suggests that the uppermost beds of the Finderup Land Nappes are Lower Palaeozoic, and perhaps equivalent to the Lower Cambrian Kap Holbæk Formation of Danmark Fjord. (3) The presence of Ordovician³ and ?Silurian fossils in the Centrum Limestone of the Sæfæxi Elv Nappe. (4) The similarities between the basal lenticular conglomerates resting on the Proterozoic sandstones and those within the Rivieradal unit.

Although none of these individually provides definite age correlations we believe that together they indicate the correlations suggested in Fig. 2.

The Finderup Land Nappes both contain sequences of laminated and cross-bedded sandstones of the Kap Holbæk Formation (with *Skolithos*) above the stromatolitic dolomites of the Fyns Sø Formation. The derivation of these nappes from the south-east suggests that early sliding occurred from an area of uplift to the east of southern Kronprins Christian Land. The folding associated with the northwesterly emplacement of the Finderup Land Nappes has resulted in an extensive area (including both allochthonous and autochthonous rocks) of northern Kronprins Christian Land having strikes at variance with the rest of East Greenland and Kronprins Christian Land in particular.

We agree with Fränkl^{2,3} that the other large nappes are derived from areas east of Kronprins Christian Land. The earlier Sæfæxi Elv Nappe mainly consists of a deep water facies of the Ordovician to ?Silurian Centrum Limestone, containing about 400 m of thin-bedded black limestones and shales. The Sæfæxi Elv Nappe which underlies the Vandredalen Nappe, consists of entirely younger strata. Both nappes have apparently been derived from a source area east of the present day Proterozoic sandstone outcrop (Fig. 1), and both have deeper water facies than equivalent beds on the platform to the west.

The oldest beds on the Vandredalen Nappe are here all assigned to the Rivieradal unit of Fränkl^{2,3}. He named several phyllite, greywacke and sandstone units and considered several nappes were present, but we believe that all the late Precambrian beds below the Campanuladal Formation on this nappe (Fig. 2) are part of a single turbidite and mud sequence on a single nappe which contains several areas (up to ~20 km²) of low-angle isoclinal folds. Because of its tectonic deformation it is not possible to estimate accurate thicknesses, but the lower part of the sequence comprises at least 1,000 m of mudstones (now slates and phyllites) and thin-bedded fine sand turbidites (some rich in calcareous detritus); and the upper part (about 1,500 m) includes thin- to thick-bedded, medium to coarse sand turbidites in addition to several mudstone units and two conglomerate horizons with rounded Proterozoic sandstone and dolerite clasts. It was one of these mass flow resedimented

conglomerates that Fränkl^{2,3} wrongly interpreted as a tillite. These turbidite sequences are followed by several hundred metres of shelf clastics, ~600 m of quartz-rich sandstones, multicoloured shales and red limestones of the Campanuladal Formation and finally by about 200 m of stromatolitic dolomites of the Fyns Sø Formation (Fig. 2).

The presence of Fyns Sø Formation stratigraphically above the Proterozoic sandstones in eastern Kronprins Christian Land (Fig. 2) implies that the Vandredalen Nappe, which also includes the Fyns Sø Formation, must have been derived from east of the Proterozoic sandstone outcrops in the Prinsesse Caroline-Mathilde Alper (Fig. 1). Fränkl^{2,3} first suggested that the sediments contained on this nappe came from the east. It seems probable that the allochthonous thick sedimentary sequences were deposited east of a swell formed by the Proterozoic sandstones which were covered only by thin late Precambrian strata below the Fyns Sø Formation. The conglomerate on the Proterozoic sandstone swell indicates local uplift of the Proterozoic sandstones and this swell may have been the source of the similar rounded clasts in the resedimented conglomerates in the upper parts of the Rivieradal unit.

If, as seems probable, the Finderup Land, Vandredalen and Sæfæxi Elv Nappes are all derived from the margins of the Iapetus Ocean¹³, they would probably have originated east of the gneiss outcrops in eastern Kronprins Christian Land. A southeasterly source for the Finderup Land Nappes would imply a total transport distance of over 150 km, whereas an easterly source for the Vandredalen and Sæfæxi Elv Nappes suggests transport of over 100 km.

The youngest age for the Silurian Profilfeldet Shale and Turbidites, which were deformed by the overriding nappes, gives a maximum age limit for the Caledonian Orogeny in East Greenland. A precise minimum age of the thrusting in Kronprins Christian Land cannot be established as no late Silurian or Devonian rocks are known. The Profilfeldet Shale and Turbidites in Kronprins Christian Land may extend into late Wenlock time; the same turbidites in Peary Land extend into the middle Wenlock¹⁴⁻¹⁶. The Profilfeldet Shale and Turbidites of Kronprins Christian Land and the Peary Land turbidites are mainly derived from the east. In both regions the turbidites succeed a prolonged episode of Ordovician and Lower Silurian platform carbonates¹⁶.

The Caledonian Orogeny in Norway occurs at the same time as in eastern North Greenland. Wenlock rocks are involved in eastward-directed thrusting near Trondheim¹⁷ and Pridoli (Downtonian) conglomerates near Bergen are post-orogenic¹⁸. It seems likely that the Caledonian Orogeny in both Greenland and Norway was due to continental collision in the late Wenlock and Ludlow, closing these northern parts of the Iapetus Ocean. It is also possible that just before collision some local areas of uplift provided a source area for the extensive Silurian turbidites of North Greenland¹⁶ in the late Llandovery and early Wenlock.

We thank N. Henriksen, H. F. Jepsen, F. Kalsbeek and J. S. Peel for critically reading the manuscript. This letter is published with the permission of the director of the Geological Survey of Greenland.

Received 18 December 1980; accepted 18 February 1981.

- Nielsen, E. *Meddr Grønland* **126**, 1-35 (1941).
- Fränkl, E. *Meddr Grønland* **116**, 1-85 (1954).
- Fränkl, E. *Meddr Grønland* **103**, 1-35 (1955).
- Adams, P. J. & Cowie, J. W. *Meddr Grønland* **111**, 1-24 (1953).
- Haller, J. *Meddr Grønland* **171**, 1-286 (1950).
- Haller, J. *Geology of the East Greenland Caledonides* (Interscience, London, 1971).
- Peel, J. S. *Rapp. Grøn. geol. Unders.* **99**, 61-72 (1980).
- Christie, R. L. & Peel, J. S. *Rapp. Grøn. geol. Unders.* **82**, 1-48 (1977).
- Troelsen, J. C. *Meddr Grønland* **149**, 1-29 (1949).
- Collinson, J. D. *Rapp. Grøn. geol. Unders.* **99**, 7-23 (1980).
- Jepsen, H. F. & Kalsbeek, F. *Rapp. Grøn. geol. Unders.* **88**, 11-14 (1979).
- Larsen, O. & Graff-Pedersen, P. *Rapp. Grøn. geol. Unders.* **99**, 111-118 (1980).
- Harland, W. B. & Gayer, R. A. *Geol. Mag.* **109**, 289-314 (1972).
- Hurst, J. M. & Surlyk, F. *Rapp. Grøn. geol. Unders.* **99**, 73-78 (1980).
- Lane, P. D. *Palaeontology* **15**, 336-364 (1972).
- Surlyk, F., Hurst, J. M. & Bjerreskov, M. *Nature* **286**, 800-803 (1980).
- Gee, D. *Tectonophysics* **47**, 393-419 (1978).
- Nicholson, R. in Harris, A. L., Holland, C. H. & Leake, B. E. (eds) *The Caledonides of the British Isles—Reviewed*, 3-18 (Geological Society, London, 1979).

Scale of body pattern adjusts to available cell number in amphibian embryos

Jonathan Cooke

National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

In many embryos, the removal of cells whose descendants would normally have formed entire parts of the body pattern is followed by apparently normal morphogenesis, which implies an ordered readjustment of the activities of the remaining cells before their potentialities become restricted. Special cell lineages cannot underlie the generation and regulation of pattern in such embryos^{1,2}. It is proposed instead that there must be some regulatory communication system in the developing embryo that normally ensures an appropriate spatial pattern of differentiation but which is also able to adjust to the removal, addition or transposition of material at a sufficiently early stage³⁻⁶. Precise models for such a mechanism have recently been suggested⁶⁻¹⁰, and have been tested experimentally¹⁰⁻¹⁴. I have performed surgical manipulations at pre-gastrula embryonic stages in two distantly related amphibian types, *Xenopus* and *Ambystoma*, and report here an assessment of the regulation achieved in terms of pattern proportions. The results are problematical for most current theories of pattern control.

Figure 1a depicts the commitment of cells to form four elements in the medio-lateral dimension of primary mesodermal pattern, during gastrulation and neurulation. Notochord is situated middorsally, and the pattern progresses laterally through somite and pro-nephros to lateral plate for the ventrally centred remainder of the cylindrical cell sheet. The pattern is probably set up in a dorso-ventral time sequence between the onset of gastrulation and the early tailbud stage¹⁵ (that is, 10–30 h, depending on species and temperature), during which about a doubling of cell number occurs along the dimension concerned². About 10⁴ cells are involved, and pattern forms across some 1–2 mm (60–100 cells) from dorsal to ventral midlines.

Figure 1b describes the first experimental challenge to the regulatory system, the manufacture of artificially small (few-celled) blastulae that develop into early larval stages without compensatory change in mitotic schedule among the remaining cells. Fate mapping¹⁶ indicates that removed blastomeres would normally have contributed pattern extending from the ventral midline, sometimes to include the entire ventral pair of elements. Figure 1c shows transverse sections of normal and miniature sibling embryos at the late tailbud stage, at which the differentiating pattern was assayed. Embryos were scanned for cell number incorporated into each pattern part, between set positions in the antero-posterior axis of pattern, that is, the anterior end of the pro-nephros and the posterior gut opening as in Fig. 1d. Scans of about 25% of this cell population (using regular subsets of the serial transverse sections and an Abercrombie correction factor) revealed that morphologically small embryos contained 25–60% of the mesodermal cell

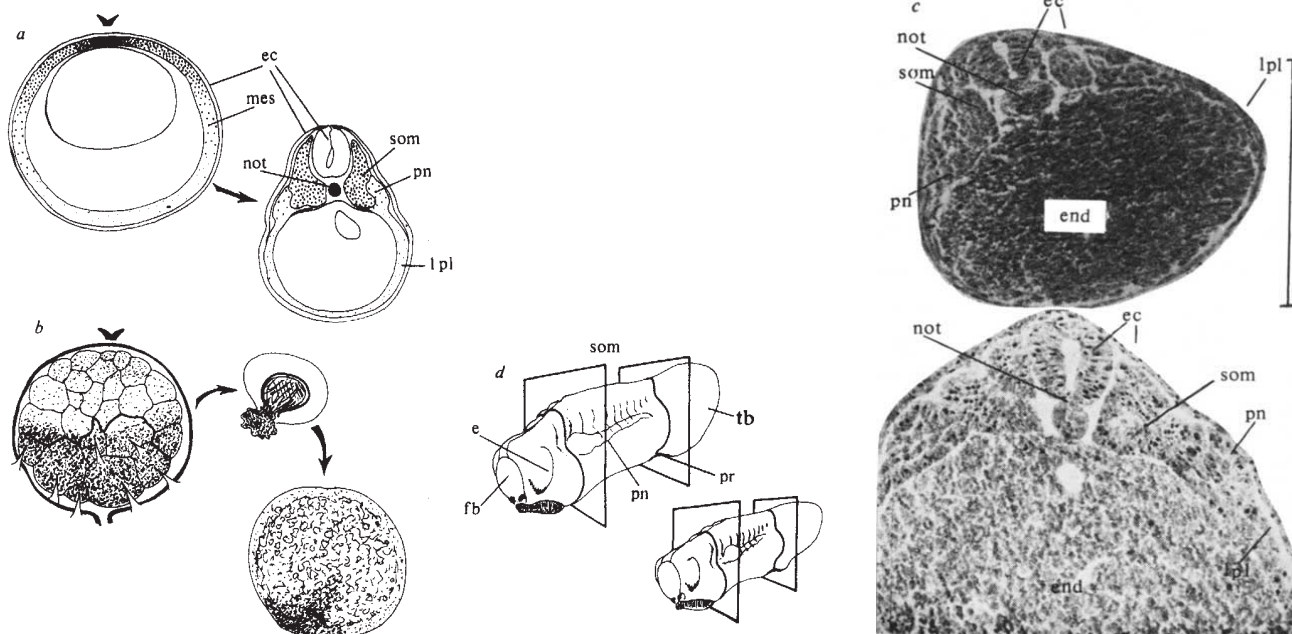


Fig. 1 a, The determination of pattern. Density of stippling in the middle mesodermal layer of the gastrula on the left, in transverse section, shows the classically recognized distribution of fates for the cells (in normal development), as they migrate forward (at right angles to the plane of the page) in the form of a coherent, roughly cylindrical, one cell-thick sheet. On the right, the same stippling represents the determined and anatomically visible zones of mesodermal structure at later (post-neurula) stages. Although these structures are solid due to piling up of cells towards the dorsal midline, their proportions in terms of cell numbers probably represent determinative events occurring in the earlier, essentially monolayered situation. Notochord (not) is a small (4%) rod-shaped mid-dorsal population, while paired somite (som) and mid-ventral lateral plate (lpl) zones each comprise ~40% of the cells. The remainder is devoted to the paired pro-nephric (pn) zones lateral to the somites. Determination seems to proceed in a dorso-ventral time sequence, starting with notochord, near the onset of gastrulation, but interactions that accompany grafting, or the co-culture of isolates, can alter development from more ventral regions until neurula stages. ec, ectoderm and neural structures induced from these; mes, mesoderm. b, Size reduction; ventral blastomeres were destroyed by pricking with a fine tungsten needle in an animal-to-vegetal sector, at morula or very large-celled blastula stages, with vitelline membrane and a thin jelly coat remaining on the embryo. Wax-coated dishes were used, containing 20% Niu-Twitty solution adjusted to pH 7.2 with dilute HCl. Dying material loses adhesivity for the healthy blastomeres and is squeezed out through an appropriately positioned slit in the coats, with rounding up and healing of the reduced blastula. Such blastulae were demembranated and kept with normal siblings in 10% Niu-Twitty on glass, until stages of fixation. The healed embryo, which reveals the loss of material in its size and reduced area of dark pigmentation, is shown to scale with the early gastrula section (in the appropriate plane) of a. ∇ marks mid-dorsal in a and b. c, Transverse section of small and normal sibling *Ambystoma* embryos at mid-trunk levels, as scanned in cell number sampling. Counts of nuclear density in the various mesodermal zones in this and the horizontal planes of section reveal that the adaptation of pattern to scale of the tissues is a genuine reflection of the participation of fewer, normal-sized cells in each structure. end, yolk; endodermal mass. Scale bar, 1 mm. d, The sampling method. A control and an operated embryo are shown, with sectional levels that bounded the region of the axis used to assay medio-lateral pattern proportions. Sampling thus scanned the mesoderm in a uniform way in each embryo but excluded the head and heart region anteriorly and the tailbud posteriorly, where relatively small proportions of the whole population are involved in morphologically complex situations. Scale regulation definitely occurs in the longitudinal aspect of head and body pattern²⁴. e, eye rudiment; fb, forebrain region; pn, outline of pro-nephros; som, outline of somites; tb, tailbud; pr, proctodaeum or original blastopore.

Table 1 Proportions in experimentally small embryos, in the patterns of original polarity in doubly organized embryos, and in control siblings

Species	Sibling set	% Of control mesodermal cell number	% Of control number of cells traversed in cross-section for small embryos, or % of host cross-sectional population involved for double dorsals	% Mesoderm cells devoted to:				
				Notochord	Somite	Pro-nephros	Lateral plate	
<i>Ambystoma</i> small embryos	1 Control	51	69	3.7	44	10	42	
	Exp.			4.5	44	12	39	
	2 Control	{ 36 25	58 35	4.2	39	11	46	
	Exp.			5.5	40	14	40	
				5.0	39	ND	56	
				4.1	45	8.7	42	
	3 Control*	{ 41	59	3.7	38	11	47	
	Exp.			4.8	39	11	45	
Double-dorsal embryos				1 Control	3.8	50	10	36
				Exp.	4.1	46	11	39
	4.2	43	9.4		43			
	4.6	48	12		35			
	56	6.0	51		8.7	34		
	2 Control	{	74	3.9	37	10	49	
	Exp.			4.3	43	12	41	
				64	4.1	36	9.5	50
59				5.0	36	15	43	
<i>Xenopus</i> small embryos	1 Control*	{	68	4.2	43	6.8	46	
Exp.	4.0			41	9.0	46		
	4.0			37	11	48		
	5.2			46	6.8	42		
	2 Control*	{ 61 39	72	4.0	39	8.0	49	
Exp.	4.0			33	10	53		
	4.5			40	1.5	54		
	4.0			44	11	41		
	3 Control	{ 30	61	4.0	39	8.0	49	
Double-dorsal embryos	Exp.			4.0	42	10	44	
	42	60	4.1	45	7.0	44		
	1 Control	{	58	4.0	37	9.2	50	
	Exp.			5.0	44	7.8	43	
				4.1	39	8.4	48	
				68	5.1	43	11	41
	2 Control	{	67	6.0	44	7.8	42	
	Exp.			63	4.2	40	8.4	47
63				6.7	44	9.0	40	
65				5.9	45	9.7	39	
3 Control	{	60	5.1	49	7.6	38		
Exp.								

For the method of scanning patterns, see text and Figs 1a, d and 2b. ND, not differentiated.

* Mesoderm population within controls differed over about a 20% range; the mean being used to derive % values for experimentals within each set. Two significant figures are used throughout, notochord and pro-nephros being small structures, somite and lateral plate large.

number in synchronous control siblings, leading to patterns in which 30–70% of the normal numbers of cells were transected per section in the pattern dimension being considered. Structures form essentially in paired strips or zones between notochord and lateral plate, so that the results of pattern regulation can be expressed as percentages of cells devoted to each structure, relative to those in control siblings.

Cell numbers and the proportions devoted to each pattern element can be assessed in the same way after a second type of operation, the grafting of a dorsal blastoporal lip or 'organizer' region from a donor early gastrula to a site in the ventral marginal zone of a late blastula host^{13,17} (Fig. 2a). In these circumstances the embryo differentiates mesodermal patterns centred round two typical dorsal midline structures, but with full cellular continuity at their lateral interface or plane of mirror symmetry within the original cylinder of the mesodermal mantle. This is because the grafted organizer, while adhering to its own presumptive fate in development as the dorsal midline, has also interacted during gastrulation with the host mesodermal cell population to organize a second axial pattern, embracing ~40% of that population and centred on itself. Cell counts as in Fig. 1d show that the total population within these mirror-symmetric mesoderms is indistinguishable from that in normal siblings at early differentiated stages as studied here (see also ref. 2). Thus, as after the previously described operation, mitotic schedule has been unaltered during cell interactions that adjust pattern.

Table 1 gives pattern proportions in embryos of every set so far examined. Because implanted organizers have only belated opportunity to interact with surrounding material, compared

with the host's original one, most models for the regulatory mechanism can explain how these may involve less than half the host tissue in a pattern of new polarity, as is usually observed. Attention is therefore confined in Table 1 to proportions in the single patterns of miniature embryos and to those in the patterns of original polarity, around the host dorsal midline, in the doubly organized ones. The behaviour within these latter patterns has the most distinctive relevance in assessing models (see discussion below). However, note that when cell numbers are pooled from both patterns for each structure, embryos with two dorsal midlines have only a slight under-representation of 'ventral' parts compared with normal, whereas an obliteration of at least the lateral plate territory might have been expected due to involvement of its tissue in new dorsal parts.

The following points emerge from the data: (1) There is appreciable variation in proportion even among normal sibling embryos, perhaps exceeding that estimated during formation of the proximo-distal pattern of the bird limb¹⁸. Nevertheless, as populations of observations, proportions within the homologous patterns at equivalent stages are remarkably similar within these two distantly related species. (2) Regulation is highly effective in assigning appropriate numbers of the available cells to the four pattern parts. Pro-nephros is relatively the most variable in proportions, in normal development and particularly in very small patterns where it may hardly exist. An apparent tendency for lateral plate to be over-represented in cases where pro-nephros is under-represented may imply that the latter is determined after somite, in tissue which will otherwise become lateral plate. If we therefore estimate pattern as three elements only, that is, notochord, somite and pro-nephros plus lateral plate,

normal and experimentally small embryos become highly comparable populations in terms of proportions. Only notochord, probably determined soon after onset of gastrulation, tends to be rather over-represented where the scale is small. Regulation has been strikingly successful, as obliteration of lateral plate and often of pro-nephros would otherwise have resulted because of the particular cells deleted. A quantitative estimate of the degree of regulation after these experimental challenges is beyond the scope of this study due to the level of variation in the outcome of undisturbed pattern formation. (3) A new organizer, situated ventrally, exerts an active influence in scaling down the pattern formed, with normal polarity around the original dorsal midline, so that both contrapolar versions of the pattern may be almost complete despite sharing a normal-sized tissue sheet during their formation. This requires, as observed, that the pattern centred round the host organizer (as well as that around the new one) has fewer cells in its dorsal parts than do normal versions.

Recent models of control of pattern in early development have arisen mainly from the notion of positional information⁶. This postulates a cell variable or signal, describable as a quantity (for example, morphogen substance concentration), and distributed in a monotonically graded profile between set 'boundary' values across the available tissue for each pattern dimension.

Cells are held to interpret this gradient by adopting developmental pathways, or determined states corresponding to particular pattern elements, in response to experience of particular absolute ranges of the signal variable. Thus stated, the idea is a formalism independent of hypotheses as to underlying mechanisms; in Fig. 3 the formalism is used to represent the normal medio-lateral pattern (a) and those actually seen after the two experimental challenges described here (version 3 of b and c). The gradient concept agrees well with the order and polarity shown by the sequences of parts in patterns, and for the

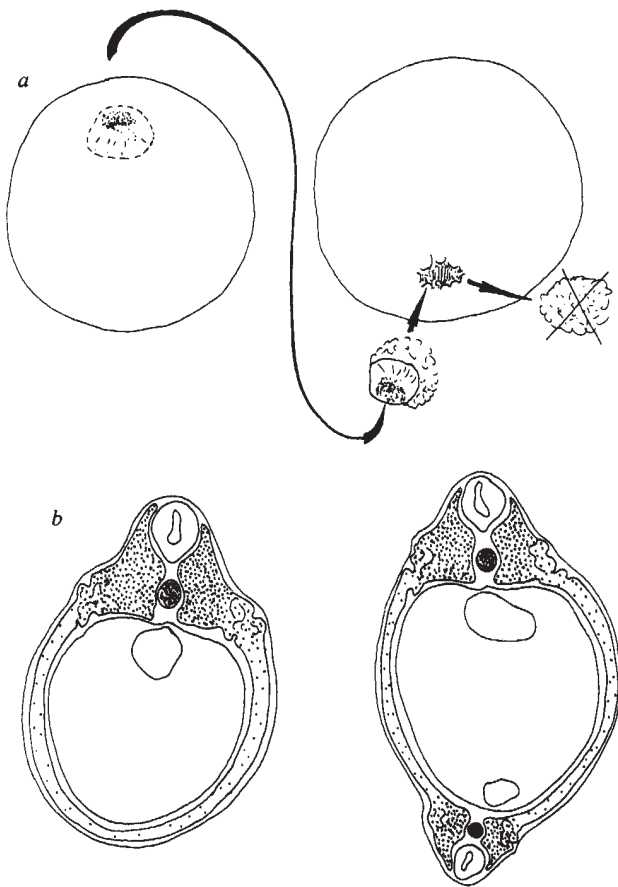


Fig. 2 a, Early gastrula donor and late blastula host viewed from vegetal aspect, presumptive dorsal midline uppermost, to show the implantation of the cell group from the dorsal blastoporal lip into a ventral marginal position at a prepared site. In the control sham operation, which is followed by normal development, the donor site is simply excised and replaced with another, similar cell group. b, Schematic transverse sections (as in Fig. 1a) through a normal embryo, and one having developed a second dorsal axis of pattern through implantation of a dorsal lip, as in a. The axial level is nearer the anterior of the two boundaries indicated in Fig. 1d. Both experimental patterns, though sharing one cell population normal for the embryo's age, have approximately normal proportions. Results in Table 1 include only proportions for the pattern retaining its original polarity in these embryos, that is, the one orientated normally in the right-hand diagram. The effect of the new organizer on the scale of structures in the pattern around the old one is the most troublesome feature for models that would explain the results on the basis of positional information.

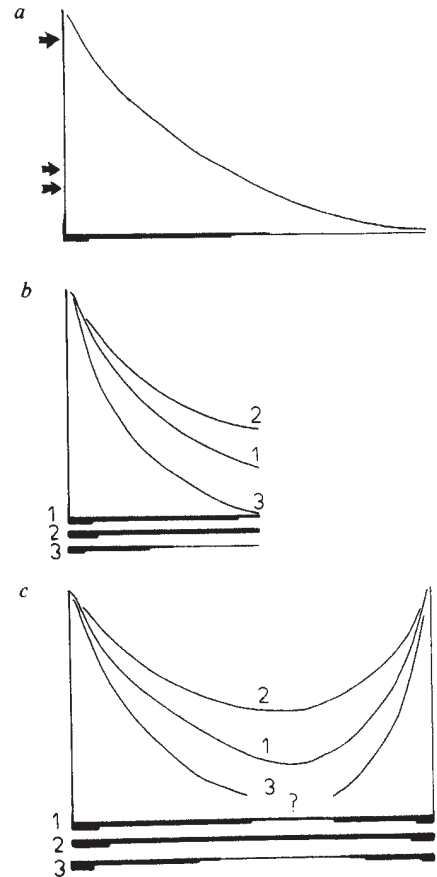


Fig. 3 a, A gradient profile which should be taken to represent formally, in scale and proportions, the normal medio-lateral pattern within the mesoderm at determination. The four thicknesses along the baseline represent the extents of zones devoted to the four pattern parts, in terms of cell number within the originally monolayered mesodermal cylinder. Arrows on the ordinate mark threshold values for a positional variable that would set the boundaries between parts on the positional information paradigm. The monotonic gradient is depicted as concavely nonlinear because most proposed mechanisms would give such a profile for, say, morphogen concentration after diffusion from a set (organizer) source. The present experimental system offers no evidence for any dynamic second organizing boundary corresponding to a local 'sink'. b, The experimental challenge of small embryonic size. Profile versions 1 and 2 represent expected results on some simple diffusion models (see text), while version 3 shows the observed result, and therefore the gradient profile that any true positional information will have assumed in order to produce it. c, The experimental challenge of double organization in a normal-sized field. Profile and baseline versions 1 and 2 again represent the range of results (from non-interaction, to mutual 'flooding out' of gradient values) expected from the simplest model mechanisms. Version 3 represents the actual behaviour and thus the necessary gradient profile if positional information is to be used. A true positional gradient for this vertebrate pattern dimension would thus have to (1) convey significant local information across much of its extent, and therefore be not extremely nonlinear (note that pro-nephros is both correctly positioned and significantly size-regulated across a range of scales, and is near the midpoint of pattern) and (2) be able to restore its lowest values to cells normally experiencing middle values, when the normal low cells were removed or replaced by a second upper boundary. These are demanding requirements, and alternative theories of pattern control might be more parsimonious, and potentially testable.

phenomenon of the organizer (that is, a pre-determined 'upper boundary' cell group controlling each pattern). But concrete models for positional information mechanisms must be able to explain, or produce during computer simulations, gradient landscapes whose scales and profiles would give the empirically seen results when interpreted as in the formalism.

Models where the organizer is a source at a pre-set level for a diffusing morphogen, and cells elsewhere are sinks, are inadequate. In such schemes, each set of parameters for diffusion, synthesis and destruction of morphogen, implies a gradient or a set time sequence of gradients that occupy particular spaces when embracing a particular concentration range. They would thus lead to profiles between versions 1 and 2 of Fig. 3*b* and *c*, corresponding to lack of regulative replacement of ventral pattern, or even to the expansion of dorsal parts due to 'flooding out' of ventral positional values. Models which use reaction-diffusion mechanisms^{8,19,20} are better, as they can produce certain size-regulated gradient profiles, and steepen gradient slopes in systems that have become bipolar in a way that would correspond with present results. Their capacities are stretched to the limits of plausibility, however, by the shape of gradient that would be required here (see Fig. 3 legend) and by the lack of evidence that any region develops 'ventral' organizer properties complementary to those of the dorsal lip. Models where each organizer acts initially to polarize morphogen transport in surrounding tissue, and thus produce its gradient, lead most easily to the necessary gradient profiles. However, there is no indication of structure or physiology that could underlie massive 'uphill' substance transport across embryos.

Alternatively, we can question the hypothesis that cells form the pattern by interpreting a positional gradient alone. Diffusible signal substances made within and specific to each developing pattern part could be involved, acting as size/proportion sensors for the system because the rate at which they are made depends on how many cells are proceeding towards a particular determination, whereas the rate at which they are destroyed depends on total tissue present. These might control the sizes of elements by feedback inhibition that switches cells towards each successive state in a series. The role of each organizer would then be simply to ensure that onset of cell determination *per se*, leading to production of the succession of specific inhibitors, occurs as a coherent (medial to lateral) wavefront across tissue. Demands for sophisticated regulation in such a gradient system would be avoided. This type of model, further described elsewhere^{21,22} makes predictions which are testable and contrast with those of positional information theory²¹, as well as accounting for interaction between organizers in scaling pattern parts (see Fig. 3*c*); see ref. 23 for discussion of formally similar models.

I thank June Colville for histology, and the MRC for supporting this work.

Received 29 January; accepted 11 March 1981.

1. Dan-Sohkawa, M. & Sato, N. *J. Embryol. exp. Morph.* **46**, 171-185 (1978).
2. Cooke, J. J. *Embryol. exp. Morph.* **51**, 165-182 (1979).
3. Driesch, H. *The Science and Philosophy of the Organism*, 2nd edn (Black, London, 1929).
4. Weiss, P. Q. *Rev. Biol.* **15**, 177-198 (1950).
5. Waddington, C. H. *J. exp. Biol.* **15**, 377-381 (1938).
6. Wolpert, L. *J. theor. Biol.* **25**, 1-48 (1969).
7. Goodwin, B. C. & Cohen, M. H. *J. theor. Biol.* **25**, 49-107 (1969).
8. Gierer, A. & Meinhardt, H. *Kybernetik* **12**, 30-39 (1972).
9. Meinhardt, H. *Rev. physiol. biochem. Pharmac.* **80**, 47-104 (1978).
10. Tickle, C., Summerbell, D. & Wolpert, L. *Nature* **254**, 199-202 (1975).
11. Wolpert, L., Clarke, M. R. B. & Hornbruch, A. *Nature* **239**, 101-105 (1972).
12. Summerbell, D. *J. Embryol. exp. Morph.* **50**, 217-233 (1979).
13. Cooke, J. J. *Embryol. exp. Morph.* **28**, 13-26, 27-46, 47-56 (1972); **30**, 283-300 (1973).
14. Herth, W. & Sander, K. *Arch. EntwMech Org.* **172**, 1-27 (1973).
15. Forman, D. & Slack, J. M. W. *Nature* **286**, 492-494 (1980).
16. Keller, R. E. *Dev. Biol.* **51**, 118-137 (1976).
17. Spemann, H. & Mangold, H. *Arch. mikrosc. Anat. EntwMech.* **100**, 599-638 (1924).
18. Summerbell, D. & Wolpert, L. *Nature* **244**, 228-229 (1973).
19. Sibatani, A. *J. theor. Biol.* **75**, 149-166 (1978).
20. Papageorgiou, S. *Biophys. Chem.* **11**, 183-198 (1980).
21. Cooke, J. *Am. Zool.* (in the press).
22. Rose, S. M. *Growth* **31**, 149-164.
23. Meinhardt, H. & Gierer, A. *J. theor. Biol.* **85**, 429-450 (1980).
24. Cooke, J. *Nature* **254**, 196-199 (1975).

Sequential functions of the bithorax complex of *Drosophila*

Gines Morata & Stephen Kerridge

Centro de Biología Molecular, Universidad Autónoma de Madrid, Canto Blanco, Madrid 34, Spain

The bithorax genes are a group of homoeotic genes of *Drosophila* whose function is related to the control of segment development. They form a gene complex in which at least seven different loci have been identified¹⁻³. Mutations at each of these loci produce a specific homoeotic transformation whereby a segment is transformed either in part or completely into another³⁻⁵. In embryos completely deficient for the bithorax genes, all the thoracic and abdominal segments resemble mesothoracic segments³. This observation, together with the segmental specificity of the mutant phenotypes, led Lewis³ to suggest a model of genetic control in which the type of development in each segment is specified by the activation of a fraction of the total number of bithorax genes. Thus, lack of activity of the bithorax genes results in mesothoracic development, activation of the bithorax (*bx*⁺) and postbithorax (*pbx*⁺) genes produces metathorax (anterior and posterior respectively), and activation of these two genes with bithoraxoid (*bxd*⁺) produces the first abdominal segment and so on. Although this model explains the genetic and developmental data, it leaves unexplained formation of the prothoracic segment. We now describe two unexpected results: deficiencies for bithorax genes can lead to segments being transformed into prothorax, and there is a temporal sequence in the function of the bithorax system.

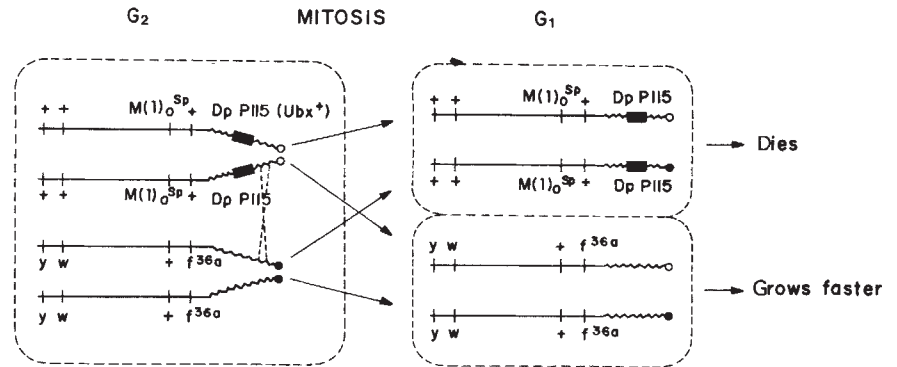
We have studied the development of mutants in the locus *Ultrabithorax* which behave as deficiencies for some of the bithorax genes^{1,2,6}. *Trans* combinations of *Ubx* alleles with any of the recessive bithorax pseudoalleles produce the phenotype of the recessive allele in question. As the *Ubx* alleles, like all bithorax deficiencies, are zygotic lethals, they cannot be studied in homozygous adults. Thus, description of the bithorax deletions has so far been restricted to the phenotype before death of embryos or first instar larvae³. Here we have used an alternative approach, whereby we induced clones of cells mutant for *Ubx* in fully viable heterozygous embryos or larvae and examined the phenotypes in the cuticle of adult flies, whose segment patterns are better known and easier to distinguish than those of the embryo or larva. Most of the zygotic lethals are cell viable in the adult cuticle⁷, and mutant clones can be recognized because they are labelled with cuticular markers (see Fig. 1). This method has the advantage that mutant clones can be produced at any desired developmental stage. In most experiments we used the *Minute* technique⁸ to obtain very large clones in which the phenotype can be clearly defined.

We chose the mutation *Ubx*¹³⁰ for most of the experiments because it produces very strong mutant phenotypes over *bx*, *pbx* and *bxd*. To avoid the possibility of effects of other mutants in the *Ubx*¹³⁰ chromosome, the flies used in the clonal analyses carried *Ubx*¹³⁰ over the deletion *Df(3R)P9*. This lethal genotype was rescued by a duplication in the first chromosome that carries all the bithorax genes in wild-type condition (see Fig. 1 legend for technical details).

Clones were induced at 4 ± 2 , 7 ± 1 , 17 ± 1 , 25 ± 1 and 58 ± 8 h after egg laying (AEL). Those induced at 4 ± 2 h AEL will be considered as blastoderm clones, and the rest post-blastoderm.

We examined blastoderm clones all over the body. In dorsal structures they differentiated normally in all segments except the metathoracic and first abdominal ones. In agreement with previous results^{9,10}, clones in the haltere showed a transformation into wing, because *Ubx*¹³⁰/*Df(3R)P9* is deficient for bithorax and postbithorax activities. For reasons discussed

Fig. 1 Details of mitotic recombination to produce clones of cells defective in bithorax activity in $M(1)o^{Sp}Dp(3;1)P115/y w f^{36a}$; $In(3LR)Ubx^{130}$, $Ubx^{130e^s}/Df(3R)P9$ flies. The lethality and mutant phenotype of the combination $Ubx^{130}/Df(3R)P9$ in the third chromosome is rescued by $Dp P115$ which carries wild-type alleles of the bithorax genes. The diagram shows a mitotic recombination event in the first chromosome (the genetic constitution of the third chromosome remains unchanged). Recombination proximal to $Dp P115$ results in the loss of the wild-type bithorax genes in one of the daughter cells. This cell is simultaneously marked with $y w$ and f^{36a} (see ref. 12) and also loses the Minute condition, which confers a growth advantage to the marked cell. The other daughter cell dies due to the presence of $M(1)o^{Sp}$ homozygously⁸. Note that mitotic recombination distal to $Dp P115$ and proximal to f^{36a} produces clones marked with $y w f^{36a}$ that remain wild type for bithorax. We estimate these are 20–25% of the total number of $y w f^{36a}$ clones (see Tables 1 and 2).



elsewhere⁴, thoracic cells cannot differentiate if they are in an abdominal system. Thus, the absence of mutant clones in the first abdominal segment indicates *bx*d transformation. Therefore, the phenotype of the $Ubx^{130}/Df(3R)P9$ clones in the dorsal parts is the one expected for the triple deficiency for *bx*, *pbx* and *bx*d.

However, clones found in the ventral thoracic structures (legs), where bristle patterns can be distinguished as pro-, meso- or metathoracic⁹, showed one important difference compared with those found dorsally—although they all respected the antero-posterior boundary line, the posterior clones of the second (IIP, mesothoracic) and third (IIIP, metathoracic) legs showed a transformation into posterior first (IP, prothoracic) leg (Fig. 2). Clones in the first leg remained untransformed (Table 1). This is the first evidence of transformation into prothorax by a mutation in the bithorax system. Note also that two different segments, mesothorax and metathorax, show the same transformation. The clones in anterior compartments behaved differently—in the first and second legs they remained untransformed but in the third leg they showed a transformation into second. This is the phenotype of mutations in the *bx* locus.

Postblastoderm clones were induced at 7 ± 1 , 17 ± 1 , 25 ± 1 and 58 ± 8 h AEL (Table 1), the last time corresponding to the middle of the second larval period. Clones were found in all body segments and their phenotypes were as described for the blastoderm clones except that at 17 h AEL and at later times IIP and IIIP clones were no longer transformed into IP. Instead, IIP clones remained as IIP and those in IIIP were transformed into IIP; that is, they showed *pbx* phenotype. At 7 ± 1 h AEL some clones showed prothoracic and others a mesothoracic pattern. Thus, the transition period between prothoracic and mesothoracic transformation is at about 7 h of development. These results have been confirmed with nine new *Ubx* alleles induced in our laboratory¹¹ and also with Ubx^1 , which, unlike Ubx^{130} , is a point mutant not associated with any chromosomal rearrangement^{1,2,12}. In this latter case we found four blastoderm clones in IIIP and two in IIP: all transformed into IP. We have also studied the legs of a sample of 32 gynandromorphs (lent by M. P. Capdevila and F. Miñana) in which the male part was mutant for Ubx^1 . The phenotype of the gynandromorph clones was the same as the blastoderm clones.

Thus clones of the same genotype, occupying the same position in the adult structure, show different mutant phenotype depending on the time of clone induction. This can be best explained if the clones are deficient for several functions which are required in a temporal order. It also suggests the existence of an unsuspected gene in the bithorax system, required for the normal development of the mesothoracic and metathoracic posterior legs.

To test whether this gene is different from *pbx* which acts in the same position, we tested whether partial duplications of the

bithorax complex can cover the posterior prothoracic transformations but not the *pbx* transformation. We used $Tp(3)bx^{100}$ (see ref. 12), a euploid chromosome where some but not all the bithorax genes (*bx*, *Cbx* and *Ubx* but not *bx*d and *pbx*) are transposed to the left arm of the third chromosome. It has a breakpoint between polytene bands 89E2–3 (ref. 12) (within the bithorax region, which extends cytologically from 89E1 to E4) and is mutant for *pbx*. By meiotic recombination we can replace the left arm of this chromosome by a normal one generating a deletion for all transposed genes. This chromosome, $Df(3R)bx^{100}$, is also mutant for *pbx* and

Table 1 Homoeotic transformations shown by $y w f^{36a}$; $Ubx^{130}/Df(3R)P9$ clones found in the posterior compartments of the legs of $M(1)o^{Sp}Dp(3;1)P115/y w f^{36a}$; $In(3LR)Ubx^{130}$, $Ubx^{130e^s}/Df(3R)P9$ females after irradiation (500 rad) at different times in development

Age at irradiation (h AEL)	No. sides	Phenotype of the clones found in		
		IP leg	IIP leg	IIIP leg
4 ± 2	1,332	IP(22)	IP(13) IIP(2)	IP(14) IIP(3)
7 ± 1	424	IP(7)	IP(3) IIP(9)	IP(1) IIP(2) IIIP(1)
17 ± 1 25 ± 1 58 ± 8	2,030	IP(36)	IIP(36)	IIP(35) IIIP(12)

Data from irradiation at 17 ± 1 , 25 ± 1 and 58 ± 8 h AEL are pooled as there was no difference between them. The actual number of clones observed is shown in parentheses. In IIP and IIIP where homoeotic transformation is seen there is always a fraction of clones that remain untransformed. These are probably the result of mitotic recombination distal to $Dp(3;1)P115$ and proximal to *forked* (see Fig. 1). Dorsal clones induced in this experiment were untransformed except in the metathoracic and first abdominal segments. Clones in the haltere were transformed into wing, anterior clones could completely reproduce the anterior wing compartment (*bx* transformation) and posterior clones could make a posterior wing compartment (*pbx* transformation). In no case did a clone cross the antero-posterior compartment boundary, either in the metathoracic where they produce a transformation or in mesothoracic segment where they have no homoeotic effect. No mutant clones were found in the first abdominal segment. This result is similar to that described⁴ for clones homozygous for Ubx^1 and for *bx*d¹ induced during the larval period, and probably indicates the transformation of first abdominal histoblast cells into metathoracic or mesothoracic cells. As a control we looked for clones in sib flies of the genotype $Dp(3;1)P115 M(1)o^{Sp}/y w f^{36a}$; $Dp(3;3)P5/In(3LR)Ubx^{130}$, Ubx^{130e^s} where a mitotic event of the type described in Fig. 1 produces $y w f^{36a}$ clones that remain wild type for the bithorax complex. None of the clones found, either in legs (72 in total) or elsewhere in the body, produced homoeotic transformation. To test the remote possibility that transformation into prothorax could be caused by the excess proliferation of M^+ clones which could interact in some unexpected manner with surrounding slowly proliferating Minute cells, we irradiated at 4 ± 2 h AEL embryos of the genotype $Dp(3;1)P115/y w f^{36a}$; $Df(3R)P9/In(3LR)Ubx^{130}$, Ubx^{130e^s} . In this case the $y w f^{36a}$ clones proliferate at the same rate as surrounding cells. Out of five clones in IIP leg, three were clearly transformed into IP and the three clones in IIIP were transformed into IP. Thus the prothoracic transformation does not result from a difference in proliferation rate.

generates an equally strong *pbx* transformation to that of *Tp(3)bx¹⁰⁰/pbx*. We expected the wild-type gene responsible for the prothoracic transformation to be missing in *Df(3R)bx¹⁰⁰* but present in *Tp(3R)bx¹⁰⁰*, in which case blastoderm clones of the genotype *Df(3R)bx¹⁰⁰/Ubx¹³⁰* would show prothoracic transformation whereas blastoderm *Tp(3R)bx¹⁰⁰/Ubx¹³⁰* clones would not. In both genotypes post-blastoderm clones should show a *pbx* phenotype.

The experiments were conducted by the same method as those described for *Df(3R)P9*, substituting *Tp(3)bx¹⁰⁰* or *Df(3R)bx¹⁰⁰* for *Df(3R)P9*. The results are shown in Table 2 and they are exactly as expected: the prothoracic transformation is absent in *Tp(3;3)bx¹⁰⁰/Ubx¹³⁰* clones but present in

Table 2 Type of homoeotic transformation shown in the posterior compartments of the legs by blastoderm clones of the genotypes indicated

Genotype	No. sides	Phenotype of blastoderm clones found in		
		IP leg	IIP leg	IIIP leg
<i>Df(3R)bx¹⁰⁰/Ubx¹³⁰</i>	910	IP(7)	IP(6) IIP(1)	IP(10) IIIP(1)
<i>Tp(3;3)bx¹⁰⁰/Ubx¹³⁰</i>	976	IP(7)	IIP(8)	IIIP(6)

Clones were induced after irradiation (500 rad) at 4 ± 2 h AEL in female embryos of genotype *y w f^{36a}/M(1)o^{5p}Dp(3;1)P115; In(3LR)Ubx¹³⁰, Ubx^{130e}/Df(3R)bx¹⁰⁰* in the first case and *y w f^{36a}/M(1)o^{5p}Dp(3;1)P115; In(3LR)Ubx¹³⁰, Ubx^{130e}/Tp(3;3)bx¹⁰⁰* in the second. Both genotypes are phenotypically wild type for bithorax. The number of clones found is shown in parentheses. Some clones in legs IIP and IIIP did not show transformation probably for the reason mentioned in Fig. 1 and Table 1. Post-blastoderm clones of both genotypes showed mesothoracic transformation.

Df(3R)bx¹⁰⁰/Ubx¹³⁰. Therefore, the gene responsible for the prothoracic transformation is present in the transposed fragment and is not associated with *pbx*. It was previously thought that lack of bithorax function results only in mesothoracic development³. However, these results show the existence of a gene of the bithorax system that, if mutant, transforms the posterior compartments of both mesothoracic and metathoracic legs into prothorax. The gene is located to the left of *pbx* and included in the fragment of the bithorax complex missing in *Df(3R)bx¹⁰⁰*. Its function is necessary for the development of the mesothoracic and metathoracic segments and its realm of action is defined by the antero-posterior compartment boundary. Like the genes *bx*, *pbx* and *bx^d*, this one also is inactivated by *Ubx* alleles. Following Lewis' nomenclature we propose to call this gene *postprothorax* (*ppx*). Its period of function differs from that of *bx* and *pbx*; *Ubx¹³⁰/Df(3R)P9* clones show prothoracic transformation only if induced before 7 h AEL. Clones of the same genotype induced at 17 h AEL develop as mesothorax. Thus, *ppx* function is needed in early development but is dispensable thereafter, unlike *bx* and *pbx* whose functions are needed until the end of the larval period⁴.

The two sets of genes seem to act sequentially: in the posterior metathoracic leg there is an early requirement for the *ppx* gene activity, which blocks prothoracic development and directs the cells to the mesothoracic pathway. A few hours later *pbx*⁺ gene function is required, preventing mesothoracic and allowing metathoracic development. Thus the final developmental route taken by the posterior metathoracic leg is specified by the combined sequential function of *ppx* and *pbx*. This suggests a temporal mechanism in the genetic process of thoracic and abdominal determination. Each segment would undergo a succession of determinative steps, each genetically implemented by the activation or inactivation of a particular homoeotic gene. The final development of each segment would be specified by a unique combination of active genes.

The *ppx* transformation affects only ventral structures (legs) whereas *bbx* (and *bx*) transforms the dorsal (halteres) as well. We cannot explain this difference but a common feature of all the homoeotic mutants producing transformation into prothorax is that this transformation is shown exclusively in the legs (for example *Polycomb* (*Pc*)¹³, *Extra sex comb* (*Scx*)¹³ and *Multiple sex comb*¹⁴).

Although our results only show evidence for a posterior prothoracic transformation, the existence of mutations like *bx* and *pbx*, that produce homologous mesothoracic transformations in anterior and posterior compartments respectively, suggests the existence of an 'anterior' prothoracic gene which, when defective, would transform the anterior meso- and metathorax into anterior prothorax. This hypothetical gene need not be in the bithorax complex.

The existence of prothoracic transformations produced by bithorax mutants apparently contradicts with results of Lewis³. He finds that embryos homozygous for *Df(3R)P9* show a transformation of all segments into mesothorax. Our explanation for this discrepancy is that ventral landmarks used by Lewis

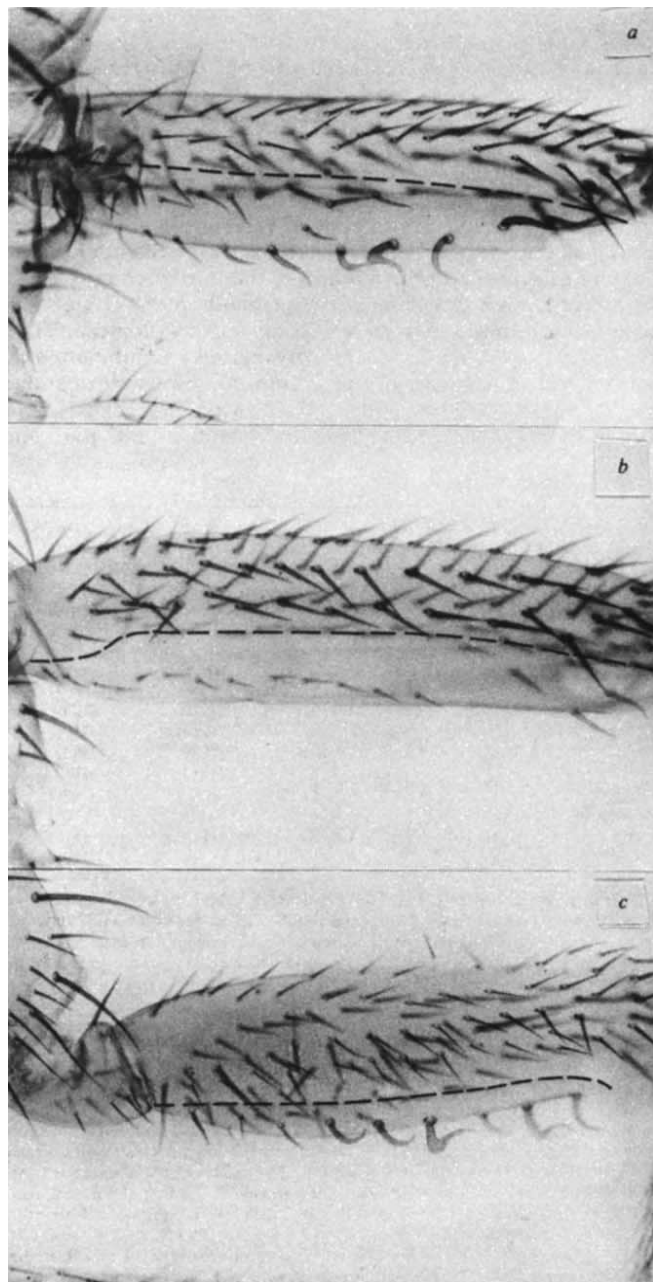


Fig. 2 Homoeotic transformation shown by *y w f^{36a}; Ubx¹³⁰/Df(3R)P9* clones (below the dotted line) in the posterior compartment of the second leg. *a*, Blastoderm clone showing transformation into posterior first leg; *b*, post-blastoderm clone not showing transformation. Note the difference in the bristle pattern of the two clones. *c* Shows a blastoderm clone in the posterior first leg for comparison.

(denticle belts) are all anterior and the prothoracic trans-formation could only be seen in the posterior compartments.

We thank Jim Kennison, Peter Lawrence, Dan Lindsley, Pedro Ripoll and Lenny Robbins for comments, Rosa González for technical help, and Victoria Lafita and Ernesto Sánchez-Herrero for typing the manuscript. S.K. is supported by an EMBO fellowship.

Received 7 November 1980; accepted 2 March 1981.

- Lewis, E. B. *Am. Zool.* **3**, 33–56 (1963).
- Lewis, E. B. in *The Role of Chromosomes in Development* (ed. Locke, M.) 231–252 (Academic, New York, 1964).
- Lewis, E. B. *Nature* **276**, 565–570 (1978).
- Morata, G. & García-Bellido, A. *Wilhelm Roux's Arch. dev. Biol.* **179**, 125–143 (1976).
- García-Bellido, A., Ripoll, P. & Morata, G. *Dev. Biol.* **48**, 132–147 (1976).
- Morata, G. & Kerridge, S. in *Development and Neurobiology of Drosophila* (eds Sidiqi, O., Babu, P., Hall, C. & Hall, J.) 141–154 (Plenum, New York, 1980).
- Ripoll, P. *Genetics* **86**, 357–376 (1977).
- Morata, G. & Ripoll, P. *Dev. Biol.* **42**, 211–221 (1975).
- Steiner, E. *Wilhelm Roux's Arch. dev. Biol.* **180**, 31–46 (1976).
- Lawrence, P. A. & Morata, G. *Dev. Biol.* **56**, 40–51 (1977).
- Kerridge, S. & Morata, G. (in preparation).
- Lindsley, D. L. & Grell, E. H. *Genetic Variations of Drosophila melanogaster* (Carnegie Institute, Washington, 1958).
- Hannah-Alava, A. *Genetics* **43**, 878–906 (1958).
- Tokunaga, C. & Stern, C. *Dev. Biol.* **11**, 50–81 (1965).

Immobilization of concanavalin A receptors during differentiation of neuroblastoma cells

Mark C. Fishman*, Paul R. Dragsten†‡
& Ilan Spector§

* Laboratory of Developmental Neurobiology, NICHD, National Institutes of Health, Bethesda, Maryland 20014, USA

† Laboratory of Theoretical Biology, NCI, National Institutes of Health, Bethesda, Maryland 20014, USA

§ Laboratory of Biochemical Genetics, NHLBI, National Institutes of Health, Bethesda, Maryland 20014, USA

Neuroblastoma cells serve as a useful model of neuronal development because compounds such as dimethyl sulphoxide (DMSO) and dibutyl cyclic AMP cause them to undergo a process of controlled differentiation in tissue culture, during which they can extend long processes, develop characteristic excitability mechanisms, synthesize neurotransmitters and form synapses^{1–3}. We have used the technique of fluorescence photobleaching recovery to study the lateral mobility of cell-surface constituents during the differentiation of neuroblastoma clone N1E-115 cells. The concanavalin A (Con A) binding sites appear as discrete patches distributed over the entire cell surface and exhibit lateral mobility in undifferentiated cells comparable with that of surface glycoproteins of other cells. After induction of differentiation, however, the vast majority of Con A binding sites become immobilized, and we present data which suggest that the mechanism of this immobilization may involve linkage to the internal actin network.

Cells of the adrenergic clone N1E-115 were grown and passaged in Dulbecco's modified Eagle's medium supplemented with 2.5% fetal bovine serum, in an atmosphere of 10% CO₂, 90% air, and replated in 35-mm plastic tissue culture dishes either with ('differentiated') or without (control) 2% DMSO. Cells were washed twice with Hank's balanced salt solution before labelling with the lectins succinyl Con A or wheat-germ agglutinin (WGA), which had been conjugated with tetramethylrhodamine isothiocyanate (E-Y Laboratories), or with the lipid analogue 3,3'-diocetadecylindocarbocyanine iodide (diI). Observation was with a Zeiss Universal microscope equipped with a ×63 1.25 n.a. water immersion lens and epifluorescence illumination.

Lateral mobility was evaluated quantitatively by the fluorescence photobleaching recovery technique^{4,5}, in which a laser beam is used as the excitation source and is sharply focused on a small area (1 µm²) of the membrane of a labelled cell. A brief intense pulse then irreversibly bleaches 40–70% of the fluorophore within that region. Monitoring the recovery of fluorescence, as unbleached molecules subsequently diffuse into the bleached region, allows calculation of the relative amounts of mobile and immobile fluorophore and the diffusion coefficient of the mobile fraction.

Con A binding sites were distributed in patches over the entire cell surface, including soma, processes and growing tips. This clustering was apparent even if the cells were prefixed with 3.7% formaldehyde, or labelled at 4 or 37 °C. No bulk movement or capping of receptors was observed. Fluorescence intensity was dramatically reduced if labelled cells were subsequently

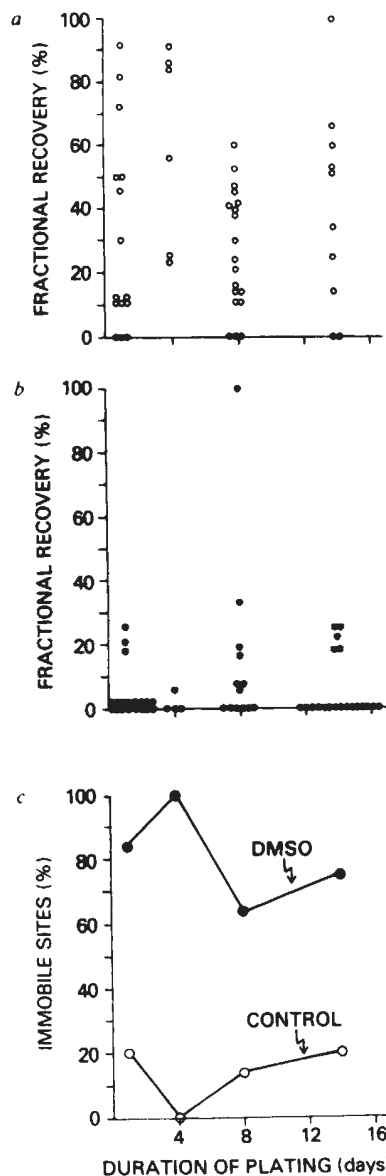


Fig. 1 Recovery of fluorescence after photobleaching. Cells were labelled with 12.5 µg ml⁻¹ rhodaminated succinyl Con A at room temperature for 20 min, and then washed twice with the balanced salt solution. Each point represents the per cent recovery of fluorescence after bleaching a site on cells incubated in control conditions (a) or after treatment with 2% DMSO (b) for the number of days indicated. c, The percentage of measurements showing immobilized Con A receptors (<10% recovery of fluorescence after photobleaching) when exposed to control (○) and differentiating (●) conditions for the number of days indicated.

‡ Present address: The Proctor and Gamble Company, Research and Development Department, PO Box 39175, Cincinnati, Ohio 45247, USA.

Table 1 Glycoprotein and lipid probe mobility on neuroblastoma cells

Probe	Treatment	Immobile sites (%)	$D \times 10^{10}$ (cm ² s ⁻¹)	Mean fractional recovery (%)
Succinyl Con A	Untreated (51)	18	1.3 ± 1.0	36 ± 28
	DMSO (all sites) (56)	79	—	7 ± 15
	Soma (31)	85	—	8 ± 20
	Hillock (13)	69	—	6 ± 8
	Process (12)	83	—	4 ± 8
	Dibutyl cyclic AMP (7)	71	—	5 ± 10
	Cytochalasin B			
	DMSO (13)	23	1.6 ± 0.8	23 ± 18
	No DMSO (10)	10	1.0 ± 0.7	28 ± 20
	Colcemid			
	DMSO (8)	75	—	10 ± 16
	No DMSO (10)	30	1.4 ± 1.0	29 ± 21
WGA	Untreated (12)	50	—	12 ± 12
	DMSO (14)	64	—	12 ± 13
	Cytochalasin B			
	DMSO (14)	50	—	14 ± 12
diI	Untreated (10)	0	101 ± 20	—
	DMSO (9)	0	122 ± 50	—

Cells were labelled with 12.5 µg ml⁻¹ succinyl Con A (10–20 min), 12.5 µg ml⁻¹ WGA (20 min) or 10 µg ml⁻¹ diI (3 min) and washed twice with the balanced salt solution. Treatment with DMSO was for 2–14 days, and with dibutyl cyclic AMP for 7 days. Cytochalasin B was used at a concentration of 10 µg ml⁻¹ during the preincubation (30 min) and left in the dish during measurements. Colcemid was used at a concentration of 0.4 µg ml⁻¹ during the preincubation (30 min) and then removed by washing. To achieve satisfactory signal-to-noise ratio, an individual diffusion measurement of diI often involved adding several successive recoveries from repeated bleaches at a single site. The immobile sites are those which showed less than 10% recovery after photobleaching. The diffusion coefficient (D) and mean fractional recovery are shown as mean ± s.d. and were calculated using the theory developed by Axelrod *et al.*⁵. D is related to the half-time for recovery, $t_{1/2}$, by $D = \gamma \omega^2 / (4t_{1/2})$ where γ is a factor that takes into account the extent of bleaching and ω is the focused laser beam radius at the e^{-2} intensity point. The fractional recovery is the per cent of fluorescence that returns after bleaching of a site. The number of sites measured under each treatment is given in parentheses.

incubated with 50 mM α -methyl mannoside, demonstrating that the Con A was binding to sites on the surface of the cells and had not been internalized.

Most Con A receptors were mobile on control cells (Fig. 1), with a diffusion coefficient of 1.3×10^{-10} cm² s⁻¹, comparable with that measured for membrane proteins in many cells⁴. In control cells neither the number of immobile sites nor their diffusion coefficient changed for periods of growth of up to 2 weeks (by which time cells had reached confluency).

However, during differentiation induced by DMSO, we found a dramatic reduction in the number of sites at which receptors were mobile (Fig. 1). This effect was seen within 36 h of the beginning of treatment and persisted throughout the 2-week study period, by which time the cells had extended long processes. There was no accompanying segregation of Con A receptors to either soma or processes, nor, as shown in Table 1, did mobile sites persist in any particular region of the cell. It is clearly possible that in conditions that leave most receptors mobile, more subtle changes in the diffusion coefficient might be found on different parts of the cell⁶. WGA receptors were also distributed over the entire cell surface but showed no evidence of a change in the number of immobile sites during differentiation (Table 1). Treatment of cells with 1 mM dibutyl cyclic AMP for 1 week also induces morphological and electrophysiological maturation², and caused a similar degree of immobilization of Con A binding sites. These effects are unlikely to be related

simply to the lowered rate of cell division seen with differentiating agents because de Laat *et al.* have shown protein diffusion to be slowest during mitosis⁷. Furthermore, serum withdrawal markedly slowed cell division⁸ and induced neurite extension but did not reduce the mean fractional recovery or increase the number of immobile sites (all sites were mobile, $n = 6$).

Further investigation suggested that the cytoskeleton contributed to the developmental immobilization. Although restrictions of mobility might conceivably derive from changes in 'fluidity' of membrane lipid^{9,10}, the diffusion coefficient of the lipid probe diI was not significantly different in control and DMSO-treated cells (Table 1), suggesting that changes in the lipid environment were not responsible. On the other hand, as shown in Table 1, dissociation of components of the cyto-



Fig. 2 Growth cone (a) and cell body (b) of a neuroblastoma cell treated with 2% DMSO, washed in balanced salt solution, labelled with succinyl Con A, fixed for 30 min in 3.7% formaldehyde and then labelled with 0.5 µg of NBD-phalloidin in 0.5 cm³ of balanced salt solution for 30 min and washed twice. It shows the correlation of the internal actin network, labelled with the NBD-phalloidin (top) and the external Con A binding sites (middle). Bright field is shown at the bottom. In other experiments cells were labelled singly with either NBD-phalloidin (after fixation) or Con A (both before and after fixation) alone and the same patterns were observed. The NBD-phalloidin did not label unfixed cells, even if incubated for 48 h. Fluorescence photomicrographs were taken using epifluorescent illumination and standard Zeiss filters for rhodamine or fluorescein (for the NBD-phalloidin). Control experiments revealed no 'bleed through': rhodaminated succinyl Con A was not visible at the NBD excitation wavelength, and the NBD was not visible at the rhodamine excitation wavelength. Scale bar, 10 µm.

skeleton reversed the developmental immobilization. Most dramatic was the effect of cytochalasin B, an agent that disrupts microfilaments, and which reduced the percentage of immobile sites in differentiated cells from 79 to 23%. Incubations with colcemid for 30 min reversibly disrupt microtubules in these cells¹¹ but only slightly increased the mobility of Con A receptors, suggesting a lesser role for microtubules in the immobilization process. Prolonged treatment (90 min) did increase the mobile fraction, but cells in these conditions had lost all processes and many floated free. WGA receptors are immobilized by cytochalasin-insensitive mechanisms on both differentiated and undifferentiated cells. Thus, there are apparently constraints in the mobility of some surface molecules which are distinct from those which act on Con A receptors.

We pursued this suggestion of cytoskeleton-membrane interaction by correlating the distribution of internal actin with that of Con A receptors. A fluorescent derivative of the mushroom toxin phalloidin has recently been demonstrated to bind actin specifically and with high affinity¹². This 7-nitrobenz-2-oxa-1,3-diazole phalloidin (NBD-phalloidin) does not seem to alter the distribution of actin filaments (by comparison with staining with anti-actin antibodies)¹³ even though the closely related toxin phalloidin can cause conversion of G- to F-actin¹⁴. The actin network, as revealed by NBD-phalloidin, is concentrated at the cell periphery and growing tips (Fig. 2). This punctate peripheral distribution was similar in control and DMSO-treated cells, and is identical to the distribution of actin described using anti-actin antibodies¹⁵. Photographs of the Con A receptors on the same cell in the same focal plane show the remarkable correlation of Con A binding sites with regions of actin accumulation. This correlation was noted on surfaces on the cell soma, along the processes and on the growth cone. WGA binding sites, on the other hand, often were not associated with actin.

Interactions between actin and cell membrane receptors are thought to underly capping in lymphocytes and other cells¹⁶⁻¹⁸ and to guide anisotropic movements on cell surfaces¹⁹. In some systems, there is evidence for linkage proteins that mediate cytoskeleton-membrane interactions²⁰. During morphological differentiation of neurones, the immobilization of selected cell-surface components, under the influence of genetic and environmental cues, may underly acquisition of structural asymmetry and could serve not only to guide directed axon outgrowth, but also to enhance anatomical registration of pre- and postsynaptic structures.

We thank L. Barak for his gift of NBD-phalloidin, A. Waggoner for his gift of dil and M. Nirenberg, R. Blumenthal and P. G. Nelson for their support and the use of their facilities.

Received 20 October 1980; accepted 11 February 1981.

- Kimhi, Y., Palfrey, C., Spector, I., Barak, Y. & Littauer, U. Z. *Proc. natn. Acad. Sci. U.S.A.* **73**, 462-466 (1976).
- Chalazonitis, A., Minna, J. D. & Nirenberg, M. *Expl Cell Res.* **105**, 269-280 (1977).
- Nelson, P., Christian, C. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **73**, 123-127 (1976).
- Elson, E. L. & Schlessinger, J. in *The Neurosciences, Fourth Study Program* (eds Schmitt, F. O. & Worden, F. G.) 691-702 (MIT Press, 1979).
- Axelrod, D., Koppel, D., Schlessinger, J. & Elson, E. *Biophys. J.* **16**, 1055-1069 (1976).
- de Laat, S. W., Van der Saag, P., Elson, E. & Schlessinger, J. *Biochem. biophys. Acta* **558**, 247-250 (1979).
- de Laat, S. W., Van der Saag, P., Elson, E. & Schlessinger, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1526-1528 (1980).
- Seeds, N. W., Gilman, A., Amano, T. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **66**, 160-167 (1970).
- Axelrod, D., Wight, A., Webb, W. & Horwitz, A. *Biochemistry* **17**, 3604-3609 (1978).
- Frye, L. D. & Edidin, M. *J. Cell Science* **7**, 319-335 (1970).
- Spiegelman, B. M., Lopata, M. A. & Kirschner, M. W. *Cell* **16**, 253-263 (1979).
- Barak, L. S., Yocum, R., Nothnagel, E. & Webb, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 980-984 (1980).
- Verderame, M., Alcorta, D., Egnor, M., Smith, K. & Pollack, R. *Proc. natn. Acad. Sci. U.S.A.* **66**, 6624-6628 (1980).
- Wieland, T. *Naturwissenschaften* **64**, 303-309 (1970).
- Marchisio, P. C., Osborn, M. & Weber, K. *J. Neurocytol.* **7**, 571-582 (1978).
- de Petris, S. in *Dynamic Aspects of Cell Surface Organization* Vol. 3 (eds Poste, G. & Nicolson, G. L.) 643-728 (North-Holland, Amsterdam, 1977).
- Karnovsky, M. J. & Unanue, E. R. *Fedn Proc.* **32**, 55-59 (1973).
- Schindler, M., Koppel, D. E. & Scheetz, M. P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1457-1461 (1980).
- Smith, B. A., Clark, W. R. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5641-5644 (1979).
- Geiger, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 4127-4131 (1980).

Role of cyclic AMP in a serotonin-evoked slow inward current in snail neurones

Philippe Deterre*, Danièle Paupardin-Tritsch*, Joël Bockaert† & Hersch M. Gerschenfeld*

* Laboratoire de Neurobiologie, Ecole Normale Supérieure, 46 Rue d'Ulm, 75230 Paris Cedex 05, France

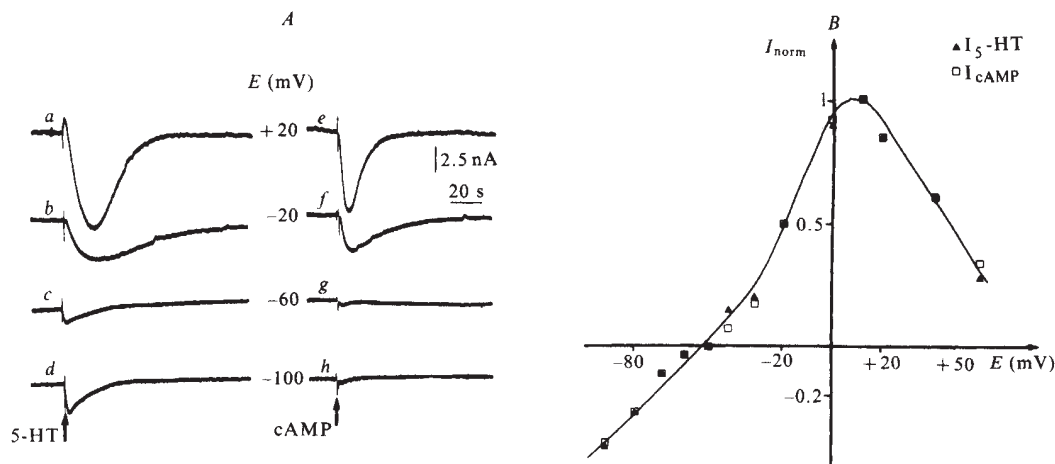
† Laboratoire de Physiologie Cellulaire, Collège de France, 11 Place Marcelin Berthelot, 75230 Paris Cedex 05, France

One model of synaptic transmission suggests¹ that transmitters modify postsynaptic permeability through the intermediary of cyclic AMP. Thus, serotonin (5-hydroxytryptamine) evokes in molluscan neurones a decrease in a voltage-dependent K^+ conductance^{2,3} which in turn generates a slow inward current when studied in steady voltage-clamp conditions³. The serotonin-induced increase of the plateau phase of the spike of an *Aplysia* sensory neurone can be mimicked by both intracellularly injected cyclic AMP and extracellularly applied phosphodiesterase inhibitors⁴, suggesting that cyclic AMP mediates the effect. We have tested whether a similar mechanism could account for the serotonin slow inward current in identified snail neurones and have found that the intracellular injection of cyclic AMP, but not of cyclic GMP or 5'-AMP, evokes a slow inward current showing similar voltage dependence, inversion potential and ionic properties to the serotonin slow inward current. Phosphodiesterase inhibitors at low concentrations (1-20 μ M) potentiate the serotonin slow inward current and at higher concentrations evoke by themselves an inward current, partially or totally occluding the serotonin and cyclic AMP currents. Finally, we have found that in homogenates of pooled identified snail neurones serotonin stimulates the adenylate cyclase, increasing its activity by 50-100%.

The experiments were carried out on identified neurones of the visero-abdominal ganglionic mass of *Helix aspersa* (cells F5, F6, E1, E2, E15, E17 of Kerkut *et al.*⁵). The ganglia were isolated in a chamber and continuously bathed in circulating saline of the following composition in mM: NaCl 120, KCl 5, CaCl₂ 6, MgCl₂ 3.5, HEPES 10 at pH 7.4. The identified neurones were impaled with double-barrelled microelectrodes connected to a voltage-clamp circuit⁶. Serotonin was applied iontophoretically or through the circulating bath solution. Tetraethylammonium (TEA) ions, Cs⁺, EGTA, cyclic AMP, cyclic GMP and 5'-AMP were injected intracellularly by passing appropriate inter-barrel currents⁷ between the barrels of double microelectrodes. One of the barrels was filled with TEA-Cl (1 M), CsCl (1 M), EGTA (0.5 M) or a 0.1 M solution of a monosodium compound of one of the nucleotides and the other with 3 M KCl. In the experiments in which both cyclic AMP and a channel blocker (or EGTA) were intracellularly injected by passing inter-barrel currents in the same cell, assemblies of three intracellular pipettes were used. NaCl was replaced in the saline by either Tris-Cl or sucrose, keeping the molarity constant. CaCl₂ in the saline was replaced by CoCl₂. When the extracellular Ca²⁺ concentration [Ca²⁺]_o was increased or TEA-Cl added to the saline, the molarity was kept constant by reducing the NaCl in the medium. Phosphodiesterase inhibitors (isobutyl methylxanthine, IBMX; RO 20-1724) were added to the circulating saline. Adenylate cyclase activity was assayed in pools of 40 identified snail neurones dissected out from 15 ganglia, according to the technique of Levitan⁸ slightly modified.

Figure 1A compares the effects of extracellular application of serotonin and the intracellular injection of cyclic AMP in the same snail neurone. When the membrane potential was held between -100 and -30 mV, serotonin evoked a biphasic inward

Fig. 1 A, Voltage-clamped snail neurone. The iontophoretic applications of serotonin (5-HT) (*a-d*) and the intracellular injections of cyclic AMP (cAMP) (*e-h*) were performed in sequence while holding the cell at different potential levels (centre). Note that the early small outward current evoked by serotonin in *a* (due to the activation of other receptors³) is not present in the cyclic AMP response (*d*). **B**, *I-V* curve relating the net currents evoked by iontophoretic applications of serotonin ($\blacktriangle$) and intracellular injections of cyclic AMP ($\square$) in a snail neurone bathed in a saline containing 30 mM TEA to suppress other components of the membrane conductance and of the serotonin response. The steady-state *I-V* curves were first obtained by stepping the voltage from a holding level (+10 mV) to different levels of potential between -90 and +50 mV and measuring the current 1 s after the onset of the pulse. The net currents evoked by serotonin and cyclic AMP were obtained by subtracting the steady-state curves obtained with serotonin and cyclic AMP from the control steady-state curves.



current (Fig. 1A, *c, d*) which, as previously shown^{3,9}, was associated with an increase in Na^+ conductance. At levels more depolarized than -30 mV (Fig. 1A, *a, b*), serotonin evoked a much slower inward current which showed a latency of 3–5 s. In contrast, no effect of intracellularly injected cyclic AMP was observed when the cell potential was held between -30 and -100 mV (Fig. 1A, *g, h*), but at potentials more depolarized than -30 mV (Fig. 1A, *e, f*) cyclic AMP also evoked a slow inward current whose amplitude increased with further

depolarization (Fig. 1A, *e*) in parallel with the serotonin response. The only difference between the two responses was that the cyclic AMP response never showed other components resulting from the activation of other serotonin receptors such as in Fig. 1A, *a*.

The current-voltage (*I-V*) curve of Fig. 1B relates the normalized values of the net currents evoked by either extracellularly applied serotonin (full triangles) or intracellularly injected cyclic AMP (squares) to the levels at which the voltage was stepped for measuring the currents. It is clear that the same *I-V* curve fits perfectly the values of the net currents evoked by either serotonin or cyclic AMP. Both currents show a maximal value at about +10 mV and an inversion potential at about -55 mV. Even if the inversion potential of both responses differed from the potassium potential, E_K (-75 mV in the snail neurones bathed in the same saline), it was sensitive to changes in the extracellular K^+ concentration. Intracellular injection of either 5'-AMP or cyclic GMP inside the same identified snail neurones had no effect.

The fact that a cyclic AMP-induced slow inward current was only seen in cells that exhibited a serotonin slow inward current, together with the similarity of the *I-V* curves of the two slow inward currents, suggests that these currents reflect the same conductance change. To test this hypothesis we compared the ionic sensitivity and pharmacology of the two currents. Neither the serotonin response nor the cyclic AMP response was affected by the replacement of two-thirds of the extracellular NaCl by either Tris-Cl or sucrose or by the presence of 30 mM TEA in the extracellular medium. The amplitude of both responses increased in parallel when $[\text{Ca}^{2+}]_o$ was increased up to 20 mM. Moreover, replacement of the $[\text{Ca}^{2+}]_o$ with Co^{2+} generated by itself an inward current and occluded the effects of both serotonin and cyclic AMP while intracellular injection of the K^+ -channel blockers TEA and Cs^+ , which evoked an inward current by themselves, also blocked the slow currents evoked by both serotonin and cyclic AMP. As was also the case for the serotonin slow inward current, EGTA intracellular injection had no blocking effect on the cyclic AMP slow inward current. Finally, when serotonin application evoked a maximal slow inward current it occluded any further effect of cyclic AMP and vice versa. Table 1 compares the properties of both slow inward currents.

These results indicate that the extracellular application of serotonin and the intracellular injection of cyclic AMP both induce slow inward currents which reflect a decrease in the same voltage-dependent and $[\text{Ca}^{2+}]_o$ -sensitive K^+ conductance.

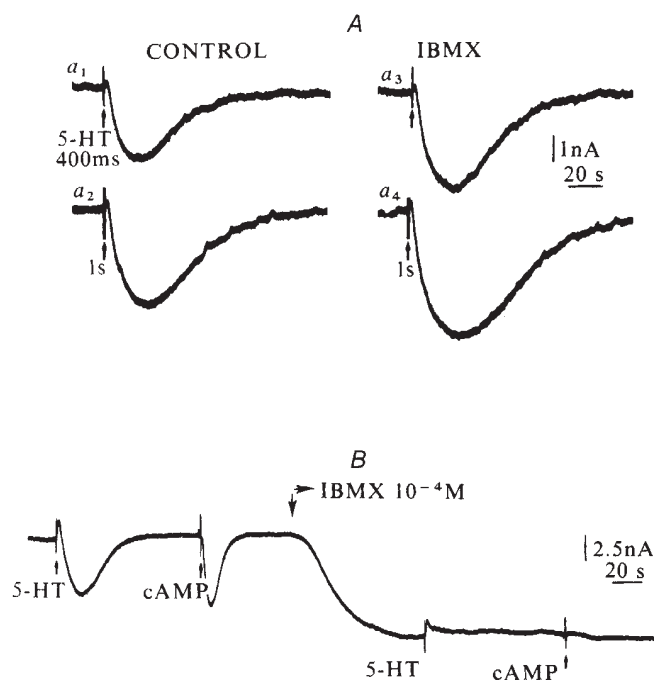


Fig. 2 A, Potentiating effect of IBMX on the serotonin slow inward current. *a*₁, *a*₂, Control serotonin slow inward currents obtained with the same iontophoretic current by varying the duration of the iontophoretic pulse *a*₃, *a*₄ Responses to the same iontophoretic serotonin pulses as in *a*₁, *a*₂ but in the presence of 20 μM IBMX in the extracellular medium. **B**, In another snail neurone, extracellular application of 100 μM IBMX evokes by itself an inward current and occludes the effects of the application of serotonin and cyclic AMP.

The next question is whether cyclic AMP mediates the effects of serotonin on the K^+ conductance whose decrease gives rise to the serotonin slow inward current. If a causal relationship exists between serotonin and cyclic AMP effects, it could be expected that the inhibitors of phosphodiesterase, the enzyme involved in the inactivation of cyclic AMP, should potentiate the response to serotonin. The results of our experiments were somewhat more complex. When IBMX was added to the bath in a concentration which, according to the cell, varied between 1 and 20 μ M, it caused, as expected, an increase in both amplitude and duration of the serotonin slow inward current (Fig. 2A, compare a_1 with a_3 and a_2 with a_4). However, at higher doses (up to 100 μ M), IBMX itself evoked an inward current which occluded both the serotonin and cyclic AMP responses (Fig. 2B). Similar effects were observed when using other methylxanthine-type phosphodiesterase inhibitors such as theophylline and with non-xanthine compounds such as RO 20-1724.

Our experiments suggest that the serotonin slow inward current results from the stimulation of an adenylate cyclase by

Table 1 Comparative properties of slow inward currents evoked by serotonin I_{S-HT} and cyclic AMP (I_{cAMP})

	I_{S-HT}	I_{cAMP}
Increase of $[Ca^{2+}]_0$	Increase	Increase
Decrease of $[Ca^{2+}]_0$	Decrease	Decrease
Co^{2+}	Block	Block
Intracellular TEA	Block	Block
Intracellular Cs	Block	Block
Intracellular EGTA	No effect	No effect

serotonin. The latency of the serotonin response and the potentiating effects of the phosphodiesterase inhibitors favour such a view because it can be assumed that there is a basal activity of the cyclase and that the higher concentrations of IBMX could evoke increased accumulation of cyclic AMP which in turn would evoke a maximal decrease in K^+ conductance and thus occlude the responses to further extracellular application of serotonin or intracellular injection of cyclic AMP.

This hypothesis was supported by assays of adenylate cyclase activity. The adenylate cyclase activity of homogenates of pooled identified snail neurones, measured by the production of cyclic AMP from ATP during 30-min periods, was 151 ± 12 fmol (mean $\pm$ s.e.m. for triplicates) in basal conditions and 283 ± 4 fmol in the presence of 10 μ M serotonin, that is, serotonin increased the enzyme activity by 87%.

In conclusion, the present results suggest that the decrease in voltage-dependent K^+ conductance associated with the slow inward current evoked by serotonin in snail neurones is actually due to serotonin stimulation of the adenylate cyclase activity. In a larger context these results, together with other effects of serotonin on molluscan neurones probably mediated by cyclic AMP^{4,10}, support the idea that some synaptic actions may also depend on the activation of a second messenger system¹.

This work was supported by grants from the CNRS, DGRST and INSERM, France. P.D. was supported by a DGRST fellowship.

Received 31 October 1980; accepted 18 February 1981.

- Greengard, P. *Nature* **260**, 101–108 (1976).
- Klein, M. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6912–6916 (1980).
- Paupardin-Tritsch, D., Deterre, P. & Gerschenfeld, H. M. *Brain Res.* (in the press).
- Klein, M. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3512–3516 (1978).
- Kerkut, G. A., Lambert, J. C., Gayton, R., Loker, J. E. & Walker, R. J. *Comp. Biochem. Physiol.* **50**, 1–25 (1975).
- Ginsborg, B. L. & Kado, R. T. J. *Physiol., Lond.* **245**, 713–725 (1975).
- Eccles, J. C., Eccles, R. M. & Ito, M. *Proc. R. Soc. B* **160**, 197–211 (1964).
- Levitan, I. B. *Brain Res.* **154**, 404–408 (1978).
- Gerschenfeld, H. M. & Paupardin-Tritsch, D. J. *Physiol., Lond.* **243**, 427–453 (1974).
- Drummond, A., Benson, J. A. & Levitan, I. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5013–5017 (1980).

Sertoli–Leydig cell communication via an LHRH-like factor

R. M. Sharpe*, H. M. Fraser*, I. Cooper* & F. F. G. Rommerts†

*MRC Reproductive Biology Unit, Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh EH3 9EW, UK

†Department of Biochemistry, Division of Chemical Endocrinology, Erasmus University, Rotterdam, The Netherlands

The primary function of the testosterone secreted by Leydig cells is the maintenance of spermatogenesis and hence fertility¹. This action of testosterone is mediated by the Sertoli cells which nourish and support the developing spermatozoa². As normal Sertoli cell function is so critically dependent on normal Leydig cell function, a regulatory influence of the Sertoli cells on the Leydig cells has been suggested^{3,4}. Indeed, follicle-stimulating hormone (FSH), which acts only on Sertoli cells⁵, can also cause profound changes in Leydig cell function^{6–8}, although how this is effected is unknown. We recently hypothesized that the Sertoli cell might be the source of a luteinizing hormone releasing hormone (LHRH)-like factor which we detected in the interstitial fluid surrounding the Leydig cells⁹. As injected LHRH agonists cause impairment of gonadal function^{10–13} and directly inhibit FSH-induced changes in Leydig cell function¹⁴ through specific membrane receptors^{15–17}, this 'LHRH-like' factor has all the correct credentials for the postulated messenger between the Sertoli and Leydig cells³. Here, we strengthen this case by demonstrating that seminiferous tubules from both the rat and the stump-tailed macaque (*Macaca arctoides*) contain a factor which has LHRH-like receptor-binding and biological activity *in vitro*, but which is immunologically distinct from native LHRH. We have also shown that this factor is secreted *in vitro* by cultured rat Sertoli cells.

Because of their ease of preparation and the specificity and affinity with which they bind radiolabelled LHRH agonists^{15–17}, isolated rat Leydig cells provide a simple, sensitive and precise radioreceptor assay for LHRH-like factors⁹. We have used this method to demonstrate that crude acid extracts of rat seminiferous tubules contain a factor which cross-reacts in parallel with unlabelled LHRH agonist (Fig. 1), whereas similar extracts of liver are inactive (data not shown). However, such crude extracts contained large amounts of protein (which may interfere in the receptor assay) and their potency decreased rapidly and progressively during storage at -20°C . An extract was therefore prepared which excluded proteins, based on methods for the extraction of native LHRH from the hypothalamus¹⁸. This acid/ethanol extract also gave parallel displacement curves to LHRH agonist in the receptor assay (Figs 1, 2), although more potent preparations tended towards non-parallelism when suppressing binding by 75% or more (see Fig. 3 and Fig. 1 legend). Acid/ethanol extracts of rat seminiferous tubules or stump-tailed macaque testis gave parallel displacement in the receptor assay and both stimulated ($P < 0.01$ and $P < 0.02$) the release of LH from rat hemipituitaries *in vitro* (Fig. 2).

Although the Sertoli cell seemed the most likely source of this LHRH-like factor, this was not certain as the seminiferous tubule preparation contained both germ cells and Sertoli cells, and was contaminated with some adherent Leydig cells. To resolve this problem we tested medium from Sertoli cell cultures. When assayed without dilution, this medium showed negligible LHRH-like activity in the receptor assay but after acid/ethanol extraction, which effectively concentrated it by a factor of 30, LHRH-like activity was clearly detected, two separate extracts giving dose-response curves parallel to LHRH agonist (Fig. 3). The specificity of this displacement can be judged by the fact that the Sertoli cell culture medium (SCCM)

extract and a rat seminiferous tubule extract gave only minor, although significant ($P < 0.05$), displacement in a Leydig cell human chorionic gonadotropin (HCG)-receptor assay which was run in parallel (Fig. 3). Interestingly, previous workers testing SCCM for inhibin-like activity reported the presence of a small molecule which stimulated pituitary LH release *in vitro*^{19,20}, and this probably corresponds to the LHRH-like factor which we have detected by receptor assay.

The nature of this LHRH-like factor produced by the Sertoli cells is unknown. It is clearly different from native LHRH, as rat seminiferous tubule extracts were largely inactive in an LHRH radioimmunoassay (Fig. 4). However, in a radioimmunoassay which we have developed using an antiserum raised against the LHRH agonist, seminiferous tubule extracts gave displacement

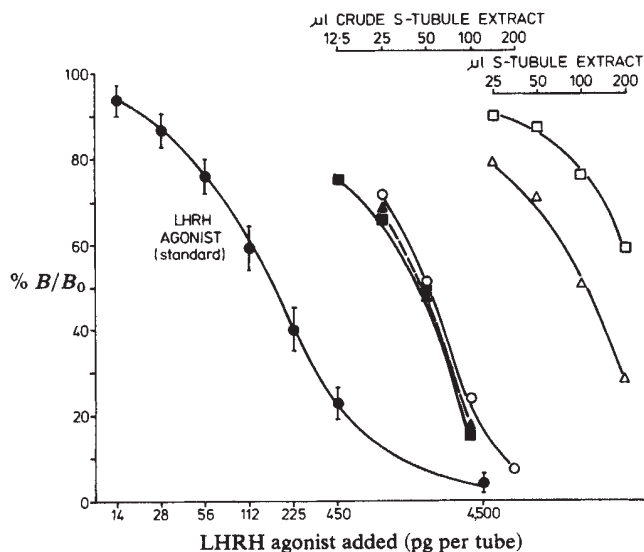


Fig. 1 Displacement of the binding of ^{125}I -LHRH agonist (D-Ser-t-bu⁶, des-Gly-NH₂¹⁰) LHRH ethylamide) to dispersed Leydig cells by unlabelled LHRH agonist and rat seminiferous tubule (S-tubule) extracts. The methods used for the preparation of Leydig cells¹⁶ and the radioreceptor assay procedure were as described previously^{9,16}, except that the volumes used were 0.15 ml dispersed cells, 0.2 ml sample/standard and 0.05 ml ^{125}I -LHRH agonist; all samples and standards were run in duplicate. To prepare crude seminiferous tubule extracts, decapsulated testes from adult Sprague-Dawley rats were dispersed with collagenase²⁹, the isolated Leydig cells removed¹⁶ and the tubules washed three times with 50 ml 0.9% saline. After the last wash the tubules were centrifuged for 10 min at 1,000g, the supernatant discarded and the precipitated tubules homogenized in 0.1 M acetic acid (0.3 ml per g tissue) using a ground glass homogenizer. The homogenate was then centrifuged for 20 min at 1,000g, the supernatant aspirated and its pH adjusted to 7.5 using 1 M NaOH. This material was then assayed at multiple doses against unlabelled LHRH agonist; dilutions of standard and test material were made using 0.1 M acetic acid adjusted to pH 7.5 and containing 0.2% bovine serum albumin (fraction V, Sigma). In subsequent studies, an acid/ethanol extraction procedure was used, based on conventional techniques for the extraction of LHRH from the hypothalamus¹⁸. Thus, washed seminiferous tubules were homogenized in 0.1 M acetic acid (1 ml per g tissue) and the supernatant collected by centrifugation. The pellet was re-homogenized in the same volume of 0.1 M acetic acid and the resultant supernatant combined with the initial extract and lyophilized. The dried powder was then extracted with 0.1 M acetic acid in ethanol (1:40, v/v) at -20°C , using 0.5 ml per g original tissue. The extract was centrifuged for 10 min at 1,000g, the supernatant aspirated and evaporated to dryness in a stream of nitrogen at 45°C . The residue was dissolved in either 0.1 M acetic acid (pH 7.5) or 0.01 M phosphate-buffered saline (pH 7.5) at a concentration of 1 ml per 10 g original tissue. The pH of the extract was then adjusted to 7.5 using 1 M NaOH and the material stored at -20°C before assay. Representative displacement curves for three crude ($\circ$, Δ , $\blacksquare$) and two acid/ethanol ($\triangle$, $\square$) extracts are shown in relation to the LHRH agonist standard curve. The latter is a composite curve (mean $\pm$ s.d.) derived from 15 successive assays and illustrates its excellent reproducibility. Parallelism of standard and sample dose-response curves was tested using two-factor analysis of variance (with replication). Of a total of 9 crude and 12 acid/ethanol extracts assayed at 3 or more dilutions, only 5 samples gave dose-response curves which showed significant ($P < 0.05$ to $P < 0.01$) non-parallelism with the LHRH agonist standard; in all 5 instances, this deviation was apparent only at high concentrations of the extract which gave more than 75% suppression of binding (for example, see Fig. 3).

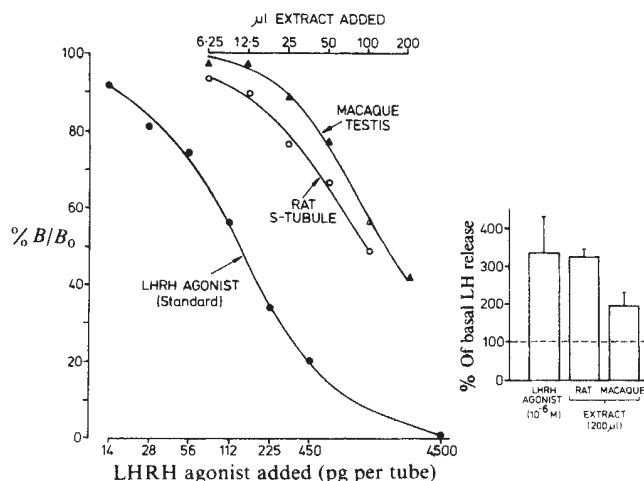


Fig. 2 *In vitro* receptor-binding activity and biological activity (inset) of acid/ethanol extracts of rat seminiferous tubules or macaque testis. Details of the extraction procedure for rat seminiferous tubules and the receptor assay method were as described in Fig. 1 legend. A single healthy, adult stump-tailed macaque of proven fertility was killed with an overdose of sodium pentobarbitone (Nembutal, May & Baker) and the testes immediately dissected out and kept on ice. After decapsulation the tissue was cut into ~ 1 -g segments which were homogenized and extracted using acid/ethanol as described for rat seminiferous tubules; the final extract contained the equivalent of 10 g tissue per ml. The *in-vitro* biological activity of the extracts was determined by their ability to stimulate LH release from rat hemipituitaries⁹ using methods described elsewhere³⁰. One half of each donor hemipituitary was incubated in the presence of the extraction medium (pH 7.5) and the contralateral hemipituitary was incubated in the presence of 0.2 ml of either the extracts or LHRH agonist (10^{-6} M). LH release in the presence of the test materials has been expressed as a % of that released by the contralateral (control) hemipituitary. LH was determined by radioimmunoassay³¹ and results are the mean $\pm$ s.d. for three incubations.

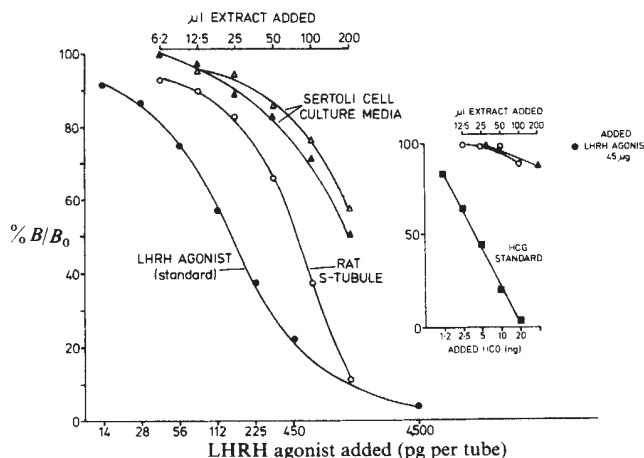


Fig. 3 Displacement of the binding of ^{125}I -LHRH agonist or ^{125}I -HCG (inset) to dispersed Leydig cells by acid/ethanol extracts of either rat seminiferous tubules or Sertoli cell culture medium. The same preparation and concentration of dispersed Leydig cells was used for both receptor assays. The HCG-receptor assay was performed as described elsewhere²⁹, and HCG CR115 (13,600 IU per mg) was used as standard. The seminiferous tubule extract was prepared as described in Fig. 1 legend, except that the pH of the initial acid extract was adjusted to 7.5, and the resulting precipitate removed by centrifugation, before lyophilization. Cultures of Sertoli cells from 21-day old rats were prepared as described elsewhere³² and maintained in culture for either 1 (Δ) or 6 ($\triangle$) days; in the latter instance the culture medium (5 ml per flask) was replaced after 1 and 5 days. Each culture flask contained ~ 700 μg protein and Sertoli and Leydig cells formed $>90\%$ and $<3\%$, respectively, as judged by morphology, histochemistry^{32,33} and LH-dependent steroid production. Culture medium was stored at -20°C before extraction. The medium from at least eight separate cultures was pooled and lyophilized after adjusting the pH to 7.5. The powder was then extracted in acid/ethanol (1 ml per 5 ml culture medium) as described for seminiferous tubules in Fig. 1 legend. The final extract contained the equivalent of 30 ml culture medium per ml.

curves parallel to the unlabelled LHRH agonist, whereas native LHRH showed <0.1% cross-reaction (Fig. 4). As there is evidence that LHRH-like molecules may be produced in other extra-hypothalamic tissues²¹, for example the ovary^{22,23}, the availability of an antibody which does not measure hypothalamic LHRH may have widespread applications.

Impairment of the secretory function of Sertoli cells, as for example in cryptorchidism^{24,25}, is associated with marked hypertrophy of the Leydig cells²⁶, and similar morphological changes have been observed locally within the testis around implants of anti-androgens³. These data have been interpreted as evidence that the Leydig cells may be under the inhibitory control of an androgen-dependent factor from the Sertoli cells³. Various findings suggest that the LHRH-like factor from the Sertoli cell, which we have described, may be this inhibitor. First, all the direct effects of injected LHRH agonists on Leydig cell function are inhibitory²¹, and it is particularly relevant that

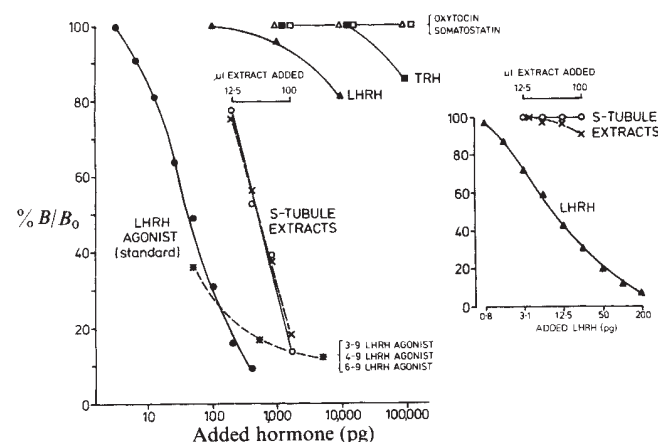


Fig. 4 Cross-reactivity of acid/ethanol seminiferous tubule extracts in radioimmunoassays for LHRH agonist and LHRH (inset). The radioimmunoassay method described previously for LHRH³⁴ was used for both assays and separation of bound and free hormone was achieved using ice-cold ethanol. ¹²⁵I-labelled LHRH or LHRH agonist was used as tracer in their respective assays and all samples and standards were run in duplicate. Details of the LHRH antibody and its specificity have been described elsewhere³⁵; LHRH agonist showed <0.1% cross-reaction in this LHRH radioimmunoassay. The antibody used in the LHRH agonist radioimmunoassay was obtained as follows. Five rabbits were immunized with ((D-Ser-t-butyl, des-Gly-NH₂)¹⁰) LHRH ethylamide (LHRH agonist) conjugated to human serum albumin using carbodiimide, as described for LHRH³⁶. Only one of the antisera (R103) cross-reacted with the seminiferous tubule extracts and this was used to develop a radioimmunoassay. The antibody was used at a final concentration of 1:12,000 and at this concentration specifically bound 26.2±4.6% (mean±s.d., n=4) of the added ¹²⁵I-LHRH agonist; the detection limit of the assay (the mass of standard required to suppress binding by 10%) was 7.5±1.1 pg. Of the peptides tested, only fragments of the LHRH agonist showed major cross-reaction with the antibody. TRH, thyrotropin releasing hormone.

the stimulatory effects of FSH on Leydig cell function (Sertoli cell-mediated effects) are totally abolished by concomitant treatment with LHRH agonists¹⁴. Second, treatment of rats with HCG, which markedly raises their intra-testicular levels of testosterone²⁷, stimulates the secretion of an LHRH-like factor into the interstitial fluid that surrounds the Leydig cells⁹, and this coincides temporally with the onset of Leydig cell down-regulation²⁸. The fact that changes as drastic as those detailed above can result from alteration of Leydig cell exposure to LHRH-like factors, suggests that this Sertoli cell hormone may have a central regulatory role in the maintenance of normal testicular function and fertility.

We thank Dr J. Sandow for helpful discussions and advice, Mrs M. Swaney for technical assistance, Drs J. Sandow and S. Dombey (Hoechst) for supplies of peptide hormones, and the National Pituitary Agency, NIAMDD, USA, for purified HCG and materials for rat LH radioimmunoassay.

Received 4 December 1980; accepted 11 February 1981.

- Hansson, V., Calandra, R., Purvis, K., Ritzén, M. & French, F. S. *Vitam. Horm.* **34**, 187-214 (1976).
- Means, A. R., Fakunding, J. L., Huckins, C., Tindall, D. J. & Vitale, R. *Recent Prog. Horm. Res.* **32**, 477-527 (1976).
- Aoki, A. & Fawcett, D. W. *Biol. Reprod.* **19**, 144-158 (1978).
- Rich, K. A., Kerr, J. B. & de Kretser, D. M. *Molec. cell. Endocr.* **13**, 123-135 (1979).
- Means, A. R. in *The Testis* Vol. 4 (eds Johnson, A. D. & Gomes, W. R.) 163-188 (Academic, New York, 1977).
- Odell, W. D. & Swerdloff, R. S. *Recent Prog. Horm. Res.* **32**, 245-288 (1976).
- Chen, Y. D. I., Payne, A. H. & Kelch, R. P. *Proc. Soc. exp. Biol. Med.* **153**, 473-475 (1976).
- Selin, L. K. & Moger, W. H. *Endocr. Res. Commun.* **4**, 171-182 (1977).
- Sharpe, R. M. & Fraser, H. M. *Nature* **287**, 642-643 (1980).
- Labrie, F. *et al. J. Androl.* **1**, 209-228 (1980).
- Rivier, C., Rivier, J. & Vale, W. *Endocrinology* **105**, 1191-1201 (1979).
- Sandow, J. & Hahn, M. *Acta endocr. copenh.* **88**, 601-610 (1978).
- Sharpe, R. M. & Fraser, H. M. *J. Reprod. Fert.* **60**, 359-368 (1980).
- Hsueh, A. J. W. & Erickson, G. F. *Nature* **281**, 66-67 (1979).
- Clayton, R. N., Katikineni, M., Chan, V., Dufau, M. L. & Catt, K. J. *Proc. natl. Acad. Sci. U.S.A.* **77**, 4459-4463 (1980).
- Sharpe, R. M. & Fraser, H. M. *Biochem. biophys. Res. Commun.* **95**, 256-262 (1980).
- Lefebvre, F. A., Reeves, J. J., Séguin, C., Massicotte, J. & Labrie, F. *Molec. cell. Endocr.* **20**, 127-134 (1981).
- Wheaton, J. E. & McCann, S. M. *Neuroendocrinology* **20**, 296-310 (1976).
- de Jong, F. H., Smith, S. D. & van der Molen, H. J. *J. Endocr.* **80**, 91-102 (1979).
- de Jong, F. H., Welschen, R., Hermans, W. P., Smith, S. D. & van der Molen, H. J. *J. Reprod. Fert. Suppl.* **26**, 47-59 (1979).
- Sharpe, R. M. *Nature* **286**, 12-14 (1980).
- Ying, S.-Y. & Guillemin, R. C. *hebd. Séanc. Acad. Sci. Paris D289*, 943-946 (1979).
- Ying, S.-Y. & Guillemin, R. *Prog. Abstr. Am. Endocrine Soc. Abstr.* **158** (1980).
- Hagenäs, L. & Ritzén, E. M. *Molec. cell. Endocr.* **4**, 25-34 (1976).
- Kerr, J. B., Rich, K. A. & de Kretser, D. M. *Biol. Reprod.* **21**, 823-838 (1979).
- Kerr, J. B., Rich, K. A. & de Kretser, D. M. *Biol. Reprod.* **20**, 409-422 (1979).
- Sharpe, R. M. *J. Reprod. Fert.* **55**, 365-371 (1979).
- Sharpe, R. M. *Biol. Reprod.* **22**, 851-858 (1980).
- Sharpe, R. M. & McNeilly, A. S. *Molec. cell. Endocr.* **18**, 75-86 (1980).
- Sharpe, R. M., Shahmanesh, M., Ellwood, M. G., Hartog, M. & Brown, P. S. *J. Endocr.* **65**, 265-273 (1975).
- Fraser, H. M. & Sandow, J. *J. Endocr.* **74**, 291-296 (1977).
- Rommerts, F. F. G., Kruger-Sewnarian, B. Ch., van Woerkom-Blik, A., Grootegoed, J. A. & van der Molen, H. J. *Molec. cell. Endocr.* **10**, 39-55 (1978).
- Rommerts, F. F. G., Kruger-Sewnarian, B. Ch., Grootegoed, J. A., de Jong, F. H. & van der Molen, H. J. *Acta. endocr. copenh.* **90**, 552-561 (1979).
- Jeffcoate, S. L., Fraser, H. M., Holland, D. T. & Gunn, A. *Acta. endocr. copenh.* **75**, 625-635 (1974).
- Jeffcoate, S. L., Holland, D. T. & Fraser, H. M. in *Physiological Effects of Immunity against Hormones* (eds Edwards, R. G. & Johnston, M. H.) 121-136 (Cambridge University Press, 1976).
- Fraser, H. M., Gunn, A., Jeffcoate, S. L. & Holland, D. T. *J. Endocr.* **61**, ix-x (1974).

Acceleration of β -receptor desensitization in combined administration of antidepressants and phenoxybenzamine

Fulton T. Crews

Department of Pharmacology and Therapeutics, University of Florida, Gainesville, Florida 32610, USA

Steven M. Paul & Frederick K. Goodwin

Clinical Psychobiology Branch, National Institute of Mental Health, Bethesda, Maryland 20205, USA

The delayed therapeutic effects of antidepressants (usually between 10 and 14 days)^{1,2} and of the tricyclic antidepressants in particular, are believed (on the basis of animal experiments³⁻⁸) to lie in a progressive decrease of the sensitivity of cortical β -adrenergic receptors. This is thought to be due to an increase in the synaptic concentration of noradrenaline, in turn accomplished by a decrease in the sensitivity of the presynaptic α_2 receptors which normally regulate noradrenaline secretion by a negative feedback mechanism. This model suggests that the desensitization of postsynaptic β -receptors by antidepressants should be accelerated by the inhibition of the presynaptic α_2 -adrenergic system, and we have indeed observed such an effect in preliminary studies with desipramine and phenoxybenzamine (PBZ) combined⁹. We now show that the administration of either tricyclic or monoamine oxidase inhibitor antidepressants in combination with PBZ, an irreversible α -adrenergic blocker^{9,10}, accelerates and intensifies the desensitization of β -adrenergic receptors. Our observations may have therapeutic implications.

Male Sprague-Dawley rats (125–150 g) were injected intraperitoneally (i.p.) with antidepressant, PBZ (7.5 mg per kg), combined antidepressant and PBZ (7.5 mg per kg), or vehicle, once daily. Rats were killed 24 h after the last dose of drug or vehicle. The cerebral cortices were dissected, weighed and homogenized using a Brinkmann polytron (setting 5, 50 s). Cortical membranes were sedimented by centrifugation (20,000g, 20 min at 4 °C), washed twice (20 vol w/v) and resuspended in 20 volumes of ice-cold 50 mM Tris-HCl buffer (pH 8.0). Aliquots of tissue equivalent to ~400–500 µg of protein were assayed for β -adrenergic receptors using ^3H -dihydroalprenolol (^3H -DHA) as described elsewhere¹¹. The binding of ^3H -DHA (0.1–3.0 nM, specific activity 32 Ci mmol⁻¹, NEN) was carried out at 23 °C for 20 min in a total incubation volume of 0.5 ml of 50 mM Tris-HCl buffer (pH 8.0). Assays were run in quintuplicate with two tubes containing 10 µM *dl*-propranolol to assess nonspecific binding. Incubations were terminated by rapid filtration at reduced pressure through Whatman GF/B glass fibre filters. Filters were washed five times with 5 ml of ice-cold buffer, dried, transferred to scintillation vials containing 10 ml of Aquasol (NEN) and counted in a liquid scintillation counter. Specific binding, defined as the total binding minus nonspecific binding, was ~85%.

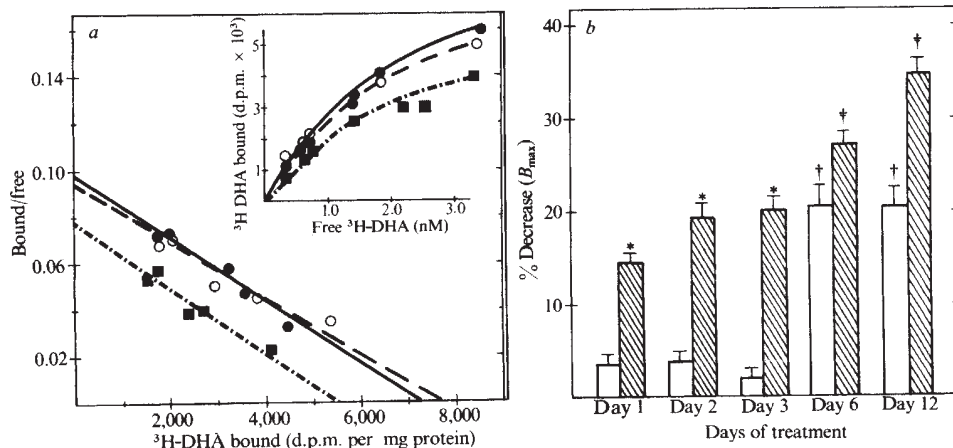
As in previous reports, acute treatment with desipramine (DMI, 7.5 mg per kg, i.p.) a prototype tricyclic antidepressant for 1, 2 or 3 days had no significant effect on the binding of ^3H -DHA to cerebral cortical membranes compared with vehicle-treated controls. Continuous administration of DMI for 6 or 12 days decreased ^3H -DHA binding by $24 \pm 4\%$ ($P < 0.05$) and $22 \pm 6\%$ ($P < 0.05$, Fig. 1), respectively. In contrast to DMI treatment alone, only one dose of combined DMI + PBZ significantly decreased ^3H -DHA binding ($14 \pm 2\%$, $P < 0.01$) compared with vehicle-treated controls (Fig. 1). Combined administration of DMI + PBZ for 2 and 3 days further reduced cerebral cortical β -adrenergic receptor density to that observed during chronic treatment with DMI alone (Fig. 1 and unpublished observations). Furthermore, during chronic administration of DMI + PBZ there was a significantly greater reduction in the density of cerebral cortical β -adrenergic receptors than that found during DMI administration alone (Fig. 1). For example, after 12 days of treatment with DMI and PBZ, specific ^3H -DHA binding was reduced by $34 \pm 3\%$ compared with $22 \pm 6\%$ with DMI alone ($P < 0.05$). All changes in ^3H -DHA binding after treatment with DMI or combined with DMI + PBZ, as determined by Scatchard analysis (Fig. 1), were due to a decrease in the density of β -adrenergic receptors and not a change in affinity (apparent K_d). These results indicate that combined DMI + PBZ administration desensitizes cortical β -receptors both more rapidly and to a greater extent than treatment with DMI alone.

Although there was a slight trend toward decreasing ^3H -DHA binding during chronic treatment with PBZ (7.5 mg per kg, i.p., data not shown) there were no statistically significant changes in ^3H -DHA binding after 1, 2, 3, 6, or 12 days of PBZ administration. The addition of PBZ (1 µM), DMI (1 µM) or combined DMI + PBZ (1 µM each) to cerebral cortical membranes *in vitro* did not significantly alter ^3H -DHA binding. As these concentrations are equivalent or even higher than those observed in washed membrane preparations, these results suggest that neither DMI nor PBZ directly interfere with ^3H -DHA binding.

To determine whether the synergistic actions of combined DMI + PBZ administration on cerebral cortical β -receptor desensitization were due to increased brain levels of desipramine, ^3H -DMI (7.5 mg per kg, i.p., 3.5 mCi mmol⁻¹, NEN) was injected into rats either alone or in combination with PBZ (7.5 mg per kg). Rats were killed 6 h later and the cerebral cortices dissected, weighed and homogenized in 20 volumes of alkaline ethanol. The radioactivity in an aliquot of the brain extract was determined in a liquid scintillation counter as described above. More than 90% of the radioactivity was extracted using this procedure. The levels of DMI and its metabolites were found to be 1.26 ± 0.09 ($n = 9$) and 1.14 ± 0.06 ($n = 9$) nmol per mg wet weight for rats treated with DMI alone or DMI in combination with PBZ, respectively. TLC of brain extracts from both groups of animals showed that 82% of radioactivity migrated with an R_f value identical to that of authentic DMI. These findings indicate that the effects of combined DMI + PBZ administration on cortical β -receptor desensitization are not due to changes in DMI metabolism.

To determine whether the rapid desensitization of cortical β -receptors found during combined PBZ + DMI (a prototype tricyclic antidepressant) treatment was true for other types of antidepressant drug, experiments were done using tranylcypromine and pargyline, two chemically different monoamine oxidase inhibitors. Treatment with tranylcypromine (10 mg per kg) alone for 1 or 2 days did not change ^3H -DHA binding. In contrast, combined tranylcypromine (10 mg per kg) + PBZ (7.5 mg per kg) treatment reduced ^3H -DHA binding by $19 \pm 2\%$ ($P < 0.01$, Fig. 2) after 1 day of treatment and $19 \pm 1\%$ ($P < 0.01$) after 2 days of treatment. Six days of tranylcypromine alone did decrease ^3H -DHA binding by $19 \pm 3\%$ ($P < 0.05$) compared with a decrease of $23 \pm 2\%$ ($P < 0.01$) with combined tranylcypromine + PBZ treatment for 6 days. Pargyline (25 mg per kg), like tranylcypromine, also had no significant effect on ^3H -DHA binding after 1 or 2 days of administration whereas after 6 days ^3H -DHA binding was reduced by $18 \pm 2\%$ ($P < 0.01$, Fig. 2). Combined pargyline (25 mg per kg) + PBZ (7.5 mg per kg) administration reduced ^3H -DHA binding by 5% after 1 day, 19% ($P < 0.01$) after 2 days and 23% ($P < 0.01$, Fig. 2)

Fig. 1 *a*, Scatchard analysis of ^3H -DHA binding to cerebral cortical membranes from rats treated with vehicle, DMI (7.5 mg per kg), or DMI (7.5 mg per kg) + PBZ (7.5 mg per kg) once daily for 2 days. The inset shows the saturation isotherm of ^3H -DHA binding as a function of increasing concentrations of ^3H -DHA. Bound/free, ratio of specifically bound to free ^3H -DHA. *b*, Effects of DMI (7.5 mg per kg) alone (open bars) or combined DMI (7.5 mg per kg) + PBZ (7.5 mg per kg) (hatched bars) on the development of cerebral cortical β -adrenergic receptor desensitization. Animals were administered vehicle or drugs i.p. once daily, and the number of β -adrenergic receptors determined 24 h after the last dose as described in the text. Values are the $\bar{x} \pm \text{s.e.m.}$ of the per cent change in the density (B_{max} as estimated by Scatchard analysis) of ^3H -DHA binding sites in cerebral cortical membranes from drug-treated animals compared with vehicle-treated controls. Control values were ~0.16 pmol ^3H -DHA bound per mg protein. Treatment with PBZ (7.5 mg per kg) alone had no significant effect on binding. Each experiment consisted of groups of six animals and was repeated twice. * $P < 0.01$; † $P < 0.05$; ‡ $P < 0.001$ (Student's *t*-test).



Values are the $\bar{x} \pm \text{s.e.m.}$ of the per cent change in the density (B_{max} as estimated by Scatchard analysis) of ^3H -DHA binding sites in cerebral cortical membranes from drug-treated animals compared with vehicle-treated controls. Control values were ~0.16 pmol ^3H -DHA bound per mg protein. Treatment with PBZ (7.5 mg per kg) alone had no significant effect on binding. Each experiment consisted of groups of six animals and was repeated twice. * $P < 0.01$; † $P < 0.05$; ‡ $P < 0.001$ (Student's *t*-test).

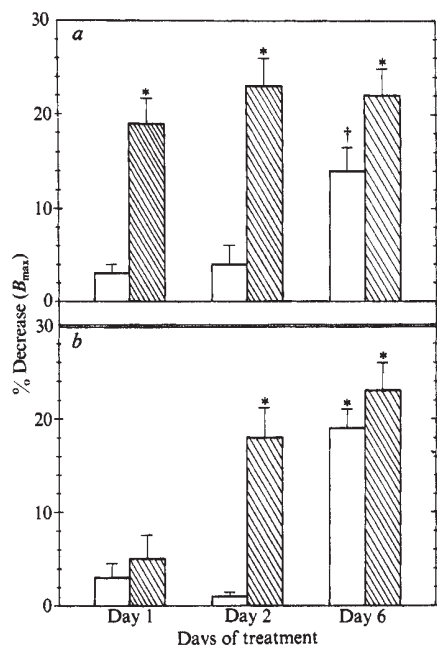


Fig. 2 *a*, Effects of tranylcypromine (10 mg per kg) alone (open bars) or combined tranylcypromine (10 mg per kg)+PBZ (7.5 mg per kg) (hatched bars) on the development of cerebral cortical β -receptor desensitization. *b*, Effects of pargyline (25 mg per kg) alone (open bars) or combined pargyline (25 mg per kg)+PBZ (7.5 mg per kg) (hatched bars) on the development of cerebral cortical β -receptor desensitization. Values are the $\bar{x} \pm s.e.m.$ of the per cent change in the density (B_{max} as estimated by Scatchard analysis) of 3H -DHA binding sites in cerebral cortical membranes from drug-treated animals compared with vehicle-treated controls. Each experiment consisted of groups of six animals and was repeated twice. * $P < 0.01$; † $P < 0.05$ (Student's *t*-test).

after 6. In both these studies treatment with vehicle or PBZ (7.5 mg per kg) alone had no significant effect on 3H -DHA binding. Thus, combined treatment of antidepressant monoamine oxidase inhibitors and PBZ can rapidly decrease β -adrenergic receptor density.

As the desensitization of cortical β -receptors during chronic tricyclic antidepressant treatment can be prevented by destruction of ascending noradrenergic neurones from locus coeruleus⁷, chronic treatment with tricyclic antidepressants probably increases the stimulation of β -receptors by increasing the synaptic concentration of noradrenaline. However, acute antidepressant administration decreases the adrenergic activity of locus coeruleus neurones apparently by increasing presynaptic α_2 -receptor inhibition¹². Recent studies in both the peripheral and central nervous systems have demonstrated that chronic antidepressant administration reduces this presynaptic α_2 -receptor inhibition^{8,13}, resulting in enhanced noradrenaline release which would be expected to desensitize postsynaptic cerebral cortical β -receptors. Thus, the increase in presynaptic α_2 -receptor inhibition during acute antidepressant treatment is opposite to the chronic and apparently therapeutically relevant effects of these agents.

In our experiments administration of PBZ probably blocks presynaptic α_2 -receptor feedback inhibition, allowing DMI, an inhibitor of noradrenaline uptake, to increase acutely the synaptic noradrenaline concentration and thereby rapidly desensitize cortical β -receptors. Consistent with this interpretation is the finding that yohimbine, another presynaptic α_2 blocker, in combination with DMI (5 mg per kg) decreased cortical β -receptor density in 4 days whereas DMI alone produced no change¹⁴.

The synergistic actions of tricyclic and monoamine oxidase inhibitor antidepressants with PBZ, an α -adrenergic antagonist, are probably due to increases in the synaptic concentrations of

noradrenaline. Rapid decreases in β -receptor binding sites have been shown to occur *in vitro* during exposure of cerebral cortical slices to β -adrenergic agonists¹⁵. The rapid decrease in 3H -DHA binding reported here after DMI + PBZ treatment indicates that cerebral cortical β -adrenergic receptors can also be rapidly desensitized *in vivo*. The rapid desensitization of cerebral cortical β -receptors during administration of antidepressants and α -adrenergic antagonists has several important implications. As postsynaptic cortical β -receptors are thought to be directly involved in the therapeutic actions of most, if not all, antidepressants³, combined treatment with antidepressant and a presynaptic α_2 -receptor antagonist may provide a therapy with rapid onset of action. In addition, the greater reduction of β -receptor density produced by combined antidepressant + PBZ, compared with antidepressant alone, indicates that administration of an α -adrenergic antagonist may enhance the effects of antidepressants in previously nonresponsive patients. Finally, the decrease in β -receptor density after only one dose of combined antidepressant + PBZ treatment could serve as a rapid screening procedure for new antidepressant drugs.

We thank the Merrell-National Laboratories for desipramine and Smith Kline and French Laboratories for PBZ. F.T.C. thanks J. Axelrod for providing laboratory facilities and constructive comments.

Received 1 April 1980; accepted 6 February 1981.

- Oswald, I., Brezinova, V. & Dunleavy, D. L. F. *Br. J. Psychiat.* **120**, 673–677 (1972).
- Klein, D. F. & Davis, J. M. *Diagnosis and Drug Treatment of Psychiatric Disorders* (Williams & Wilkins, Baltimore, 1969).
- Sulser, F., Vetulani, J. & Mobley, P. L. *Biochem. Pharmac.* **27**, 257–261 (1978).
- Banerjee, S. P., King, L. S., Riggi, S. J. & Chanda, S. K. *Nature* **268**, 455–456 (1977).
- Sellinger, M. M., Frayer, A. & Mendels, J. *Fedn Proc.* **28**, 592 (1979).
- Vetulani, J., Stawary, R. J., Dingell, J. V. & Sulser, F. *Naunyn-Schmiedeberg's Archs Pharmac.* **293**, 109–114 (1976).
- Wolfe, B. B., Harden, T. K., Sporn, J. R. & Molinoff, P. B. *J. Pharmac. exp. Ther.* **207**, 446–457 (1978).
- Crews, F. T. & Smith, C. B. *Science* **202**, 322–324 (1978).
- Paul, S. & Crews, F. T. *Eur. J. Pharmac.* **62**, 349–350 (1980).
- Cubeddu, L. X., Langer, S. Z. & Weiner, N. *J. Pharmac. exp. Ther.* **188**, 368–385 (1974).
- Bylund, D. B. & Snyder, S. H. *Molec. Pharmac.* **12**, 568–580 (1976).
- Nyback, H., Walters, J. R., Aghajanian, G. K. & Roth, R. H. *Eur. J. Pharmac.* **32**, 302–312 (1975).
- Svensson, T. H. & Usdin, T. *Science* **202**, 1089–1091 (1978).
- Johnson, R. W. *et al. Eur. J. Pharmac.* **67**, 123–127 (1980).
- Maggi, A., U'Prichard, D. C. & Enna, S. J. *Science* **207**, 645–647 (1980).

Isolation of a biologically active macrophage receptor for the third component of complement

Robert J. Schneider*†, Anthony Kulczycki Jr*‡, S. K. Law§ & John P. Atkinson*†

* Howard Hughes Medical Institute Laboratories and the Department of Medicine, † Division of Rheumatology and ‡ Division of Allergy and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA
§ Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA

C3b, the major cleavage fragment of the third component of complement (C3), has been demonstrated to bind to a specific receptor on various mammalian cells¹. Tissue-bound C3b receptors have also been demonstrated, most conclusively in renal glomeruli². On interaction with C3b, C3b receptors are thought to initiate cellular functions such as phagocytosis and to determine the fate of complement-fixing soluble and particulate immune complexes. We report here the isolation by affinity chromatography of a macrophage glycoprotein with an apparent molecular weight (MW) of ~64,000 that has properties expected of the C3b receptor. It is a cell-surface macromolecule (labelled with ¹²⁵I and lactoperoxidase) which, in its isolated state, retains the ability to bind both C3 and C3b.

Human C3 was purified by the method of Tack and Prahl³, and was demonstrated to be haemolytically active³ and to retain its ¹⁴C-methylamine-binding capacity⁴. Three separate C3 preparations were used in the experiments; in most of them a preparation with >95% methylamine-binding capacity was used. C3b and C3d were prepared from C3 by trypsin digestion followed by gel filtration on Sephadex G-100 (ref. 5), and C3d was further purified by ion-exchange chromatography on DEAE-Sephacel⁴. The purities of the C3, C3b and C3d preparations were estimated by polyacrylamide gel electrophoresis (PAGE)⁶ to be at least 95%. C3 and C3b were coupled to Sepharose beads with cyanogen bromide with an efficiency of 57–86% and prepared for use as immunoadsorbents as described previously^{7–9}.

Elicited rabbit alveolar macrophages were collected, radiolabelled using modifications of the lactoperoxidase method¹⁰ and solubilized at 4 °C for 5 min in borate-buffered saline (BBS) containing 1% NP40 and 2 mM phenylmethylsulphonyl fluoride (PMSF)⁷. The 30,000g supernatant⁷ of the solubilized fraction of $\sim 2 \times 10^8$ cells was incubated with 0.4 ml of C3b-Sepharose for 1 h at room temperature. This mixture was then centrifuged at 300g for 5 min and the immunoadsorbent beads resuspended in 6.0 ml of BBS containing 1% NP40 and 2 mM PMSF, and washed twice in 0.7×4 -cm glass columns with 4.0 ml of BBS containing 1%

NP40. C3b-binding molecules were subsequently eluted at room temperature with four successive 1.0-ml aliquots of 0.5 M acetic acid containing 1% NP40 into tubes with 0.375 ml of 2 M Tris, pH 8.6, containing 1% NP40 (to achieve complete, rapid neutralization). Using this procedure, we eluted 0.01–0.15% of the applied radioactivity from the immunoadsorbent columns.

When the acid eluate was analysed by SDS PAGE in rod gels in reducing conditions (1% β -mercaptoethanol), a single protein peak of MW $\sim 64,000 \pm 3,500$ (mean $\pm$ s.d. of 21 gels) was demonstrated (Fig. 1a). Purity, estimated by either determining the area under the curves or by cutting out and weighing the individual peaks, was routinely >90%. A similar peak was purified from the solubilized radiolabelled cells using C3b immunoadsorbent beads that had been prepared by the incubation of Sepharose with normal human serum¹¹, although other contaminating materials were also present (data not shown). No peak was observed if purification was attempted using beads prepared from heat-inactivated serum or using bovine serum albumin (BSA)-Sepharose. To exclude the possibility that trace contamination by IgG in our C3b preparation might allow co-purification of Fc γ receptor⁷, we preincubated the solubilized preparation with IgG-Sepharose. In three such experiments we recovered 92.5, 97.3 and 108% as much C3b receptor from the aliquot preincubated with IgG-Sepharose as from the aliquot not preincubated but isolated directly using only C3b-

Table 1 Rebinding of isolated C3b receptor to immunoadsorbents

Expt no.	Immunoadsorbent For initial purification	For rebinding	Inhibitor protein	% Rebinding	% Specific rebinding	% Inhibition of specific rebinding
1	C3b	C3b	—	64.0 $\pm$ 1.1	60.6	—
		C3	—	60.8 $\pm$ 1.2	57.4	—
		IgG	—	7.1 $\pm$ 0.8	3.7	—
		BSA	—	3.4 $\pm$ 0.2	—	—
2	C3b	C3b	—	69.8 $\pm$ 2.0	62.3	None
		C3b	3.0 mg C3	21.8 $\pm$ 2.9	14.3	77.0
		IgG	—	7.5 $\pm$ 4.3	—	—
	IgG	C3b	—	18.7 $\pm$ 5.0	—	—
		IgG	—	57.8 $\pm$ 1.0*	—	—
	C3b†	C3b	—	63.2 $\pm$ 2.0	56.1	—
		IgG	—	7.1 $\pm$ 3.7	—	—
	IgG‡	C3b	—	18.0 $\pm$ 0.8	—	—
		IgG	—	51.9 $\pm$ 0.4	—	—
3	C3	C3b	—	31.6 $\pm$ 1.3	26.1	None
		C3b	3.0 mg OVAL	30.2 $\pm$ 2.0	24.8	5.0
		C3b	3.0 mg C3	8.1 $\pm$ 0.2	2.7	89.7
		C3b	3.0 mg IgG	29.1 $\pm$ 0.4	23.7	9.2
		IgG	—	5.4 $\pm$ 0.5	—	—
4	C3b	C3b	3.0 mg OVAL	27.1 $\pm$ 3.6	22.9	None
		C3b	0.3 mg C3b§	15.7 $\pm$ 1.7	11.5	49.8
		C3b	1.0 mg C3b	10.5 $\pm$ 1.9	6.3	72.4
		C3b	3.0 mg C3b	5.1 $\pm$ 0.7	0.9	96.1
		C3b	0.3 mg C3	19.3 $\pm$ 5.7	15.2	34.0
		C3b	1.0 mg C3	12.7 $\pm$ 2.5	8.5	63.0
		C3b	3.0 mg C3	6.3 $\pm$ 2.2	2.1	91.0
		IgG	3.0 mg OVAL	4.2 $\pm$ 4.5	—	—
5	C3b	C3b	3.0 mg OVAL	46.6 $\pm$ 1.6	41.5	None
		C3b	0.6 mg C3b§	20.2 $\pm$ 1.8	15.1	63.6
		C3b	0.6 mg C3d	49.3 $\pm$ 2.6	44.2	-6.6
		IgG	3.0 mg OVAL	5.1 $\pm$ 1.3	—	—

The ligand-binding activity of the putative receptor macromolecule was determined in a manner analogous to that described for the Fc γ receptor^{7,13}. Briefly, aliquots of the acid eluates were added to plastic tubes containing 0.10 ml immunoadsorbent. Inhibitor proteins were preincubated with the isolated receptor for 30 min at room temperature. Mixtures of immunoadsorbents and receptor preparations were incubated for 90 min at room temperature. Washed beads, supernatants and washes of triplicate samples were counted in a γ counter and the per cent rebinding of labelled receptor determined. Data presented are from representative experiments. The amounts of protein coupled to immunoadsorbents (mg protein per ml beads) were: C3b, 2.4–11.6 mg ml⁻¹; C3, 1.7–11.3 mg ml⁻¹; rabbit IgG, 6.8–15.7 mg ml⁻¹; and BSA, 4.5 mg ml⁻¹. For all rebinding studies, the amount of C3 or C3b present was in excess of that required to bind all of the available receptor.

* Represents rebinding of Fc γ receptor, which demonstrates the expected MW $\sim 35,000$ –55,000 (ref. 7) in SDS-polyacrylamide gels.

† The supernatant from the IgG-Sepharose column shown in the first part of expt 2 was adsorbed on C3b-Sepharose. The ligand-binding activity of the acid eluate from this second column is shown here.

‡ The supernatant from the C3b-Sepharose column shown in the first part of expt 2 was adsorbed on IgG-Sepharose. The ligand-binding activity of the acid eluate from this second column is shown here.

§ In these dose-dependent inhibition studies, ovalbumin (OVAL) was added to achieve a total protein content of 3.0 mg.

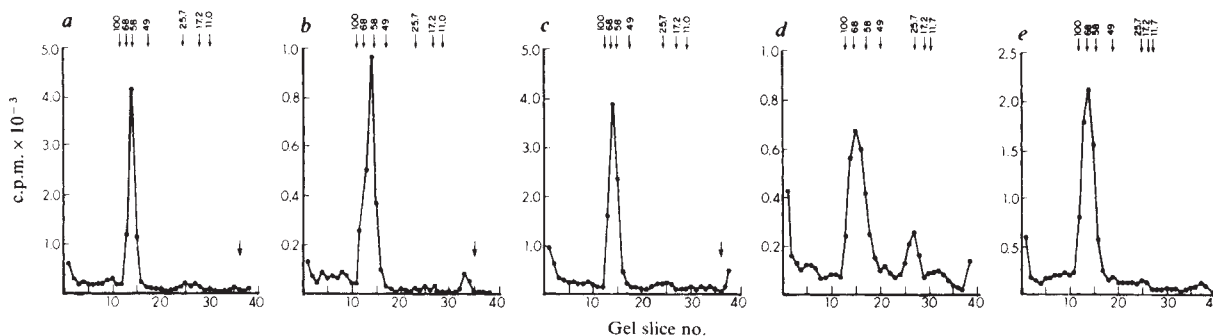


Fig. 1 SDS-PAGE in rod gels of C3b receptor. Representative results of five different types of experiment are shown. Samples were eluted with acid as described in the text and the eluates were prepared for electrophoresis using the Tris-glycine buffer system²¹ with 0.1% SDS and 9.0% acrylamide separating gels and 3.0% acrylamide stacking gels, using methods described previously⁷. Buffers for sample preparation contained 1% β -mercaptoethanol. The positions of six standard proteins are indicated by their molecular weights. Gels were sliced with an electric gel slicer, and radioactivity in 2-mm gel slices counted. *a*, Receptor from C3b-Sepharose. *b*, The ¹²⁵I-labelled, solubilized fraction was first incubated with IgG-Sepharose (to deplete Fc γ receptor from the preparation). This supernatant was subsequently incubated with C3b-Sepharose and the receptor was eluted and analysed as above. (The difference in counts between *a* and *b* reflects only the fact that one-fifth as many counts were loaded in the latter gel. As noted in the text, preincubation of the solubilized preparation with an IgG immunoabsorbent did not alter the quantity of C3b receptor obtained.) *c*, Receptor from C3-Sepharose. *d*, After collection by usual methods, the macrophages were resuspended in Hank's balanced salt solution (HBSS) at a concentration of 1.3×10^6 cells ml⁻¹. Cell suspension (8 ml) was placed on each of 45 100-mm plastic Petri dishes and incubated for 30 min at 37 °C in a 5% CO₂ incubator. The nonadherent cells were removed by washing once with 5 ml of HBSS. The plated cells were solubilized with 4 ml of BBS containing 1% NP40 and 2 mM PMSF, and freed from the plates with a rubber policeman. The plates were serially washed with an additional 4 ml of the BBS containing 1% NP40 and 2 mM PMSF⁷. The solubilized preparation thus obtained was centrifuged at 30,000g for 20 min at 4 °C and the C3b receptor isolated by affinity chromatography as above. A small peak at ~25,000 MW is present in this preparation (and also in the simultaneously prepared material obtained from the unplated cell population—not shown). Similar macromolecules were observed in 2 of the other 15 receptor preparations. *e*, Rabbit alveolar macrophages were collected and washed as above using sterile buffers and aseptic technique. Cells were subsequently suspended at 10^6 cells ml⁻¹ in sterile spinner flasks in minimal essential medium containing 10% fetal calf serum, 2 mM L-glutamine, 0.01% D-glucose, 1 mM pyruvate and 1:100 antibiotic/antimycotic mixture. Deoxyribonuclease I (10 μ g ml⁻¹) and ¹⁴C-glucosamine (1 mCi) were added to the medium. After 18–20 h culture at 37 °C, the cells were collected by centrifugation and washed twice in 0.01 M phosphate, 0.15 M NaCl, pH 7.4 (phosphate-buffered saline, PBS) containing 1% BSA followed by a final wash in PBS. Cells were solubilized at 10^8 cells ml⁻¹ in PBS containing 1% NP40 and 2 mM PMSF⁷. The suspension was then centrifuged at 30,000g for 20 min at 4 °C, and the supernatant applied to a C3b-Sepharose affinity column for 60 min at room temperature. The column was washed and eluted as described in the text to obtain receptor from intrinsically labelled cells.

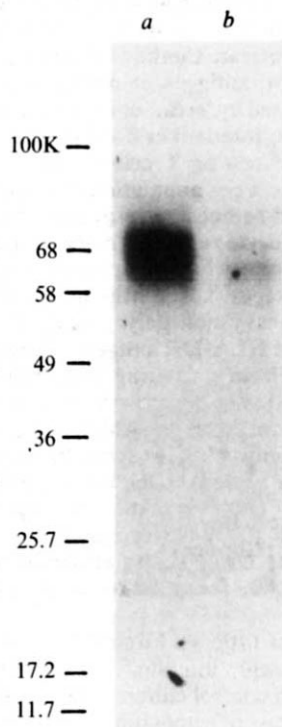


Fig. 2 SDS-PAGE in slab gels was performed according to the method of Laemmli⁶ using 0.1% SDS gels. Preparation of a linear gradient of polyacrylamide from 6 to 18% and a 3% stacking gel, gel fixation, staining, destaining and autoradiography have been described elsewhere¹³. Samples were prepared as described in Fig. 1 legend. *a*, Acid eluate from C3-Sepharose. *b*, A portion of the acid eluate in *a* was adsorbed on a new C3-Sepharose column. The rebound material was eluted with 6 M guanidine containing 1% NP40 and the sample prepared for analysis as above.

Sepharose. Macromolecules of identical MW were isolated from C3b-Sepharose and from C3-Sepharose (Fig. 1c); a similar protein was also obtained from an immunoabsorbent prepared by the coupling of C3 to Affi-Gel 10 (BioRad). The polyacrylamide gel electrophoretic mobility and pattern of the putative C3b receptor were not altered by: (1) addition of protease inhibitors (diisopropylfluorophosphate 10 mM, iodoacetamide 20 mM, or soybean trypsin inhibitor 2 mg ml⁻¹) in the solubilization buffer; (2) incubation temperature (4, 22 or 37 °C) for affinity chromatography step; (3) omission of β -mercaptoethanol from the sample buffer; or (4) presence of Ca²⁺ and Mg²⁺ during solubilization and affinity chromatography. As alveolar cells contained an average of ~80% macrophages, adherent cells containing >95% macrophages⁷ were used in three experiments to obtain C3b receptor, and an identical macromolecule was isolated (Fig. 1d). Biosynthetic labelling with ¹⁴C-glucosamine demonstrated that, like the Fc ϵ ^{9,12} and Fc γ ⁷ receptors, the C3b receptor is a glycoprotein (Fig. 1e).

The ligand-binding activity of the affinity chromatography-purified material (always lower than the purity estimated by SDS-PAGE) was determined by the ability of the acid eluates to rebinding to a new immunoabsorbent column. In 12 such rebinding experiments the specific activity (rebinding to C3b-Sepharose minus rebinding to IgG-Sepharose or BSA-Sepharose) ranged from 22.9 to 62.3% (representative experiments are presented in Table 1). The per cent of rebinding was characteristic for each receptor preparation. The variability between preparations did not reflect differences in purity but rather partial denaturation during the obligatory acid elution procedure. The ligand-binding activity of the isolated receptor was independent of Ca²⁺ and Mg²⁺; it was stable for several hours at room temperature and at least several weeks at -70 °C. The receptor rebound to both C3b-Sepharose and C3-Sepharose (Table 1, expt 1). Preincubation of the solubilized preparation with IgG-Sepharose, which removes Fc γ receptor⁷, did not alter the specific activity

of the C3b-binding protein eluted subsequently from a C3b-Sepharose column (Table 1, expt 2).

Two lines of evidence indicate that the ability of the isolated receptor to bind to both C3 and C3b was not due to conversion of C3 to either C3b or a C3b-like molecule during coupling to activated Sepharose. First, an alternative coupling procedure using Affi-Gel 10 allowed isolation of a similar macromolecule (see above). Second, both soluble C3 and soluble C3b were effective inhibitors of rebinding (Table 1, expt 3), although C3b was more effective in dose-dependent assays (Table 1, expt 4 is representative of three separate experiments). Soluble IgG, C3 and ovalbumin did not inhibit ligand-binding activity.

Aliquots of the acid eluates from the initial C3-Sepharose column were analysed in polyacrylamide gradient slab gels (Fig. 2a). Part of this material was rebound to a second C3-Sepharose column and then eluted with 6 M guanidine containing 1% NP40 (ref. 7), dialysed and analysed in gels (Fig. 2b). Electrophoretic patterns of both eluates demonstrated similar broad bands in the 58,000–68,000 MW region.

Like receptors for Fc γ ^{7,13} and Fc ϵ ^{9,12,14} previously isolated in these laboratories, the C3b-binding protein migrates as a broad band on polyacrylamide gels and retains specific ligand-binding activity. The molecular weight of the C3b receptor is higher than that of the Fc γ receptor isolated from rabbit alveolar macrophages, the latter having a MW of ~35,000–55,000 (ref. 7 and Table 1,*). Specific antibodies against analogously isolated Fc ϵ -binding molecules have been demonstrated to induce mast cell degranulation *in vitro*^{15,16}, providing definitive evidence that such binding molecules correspond to the biologically important Fc ϵ receptor. Similar evidence identifying Fc γ - or C3b-binding proteins conclusively as receptors has not yet been obtained. It now seems, however, that at least three important receptor molecules involved in immune responses are glycoproteins with apparent MWs in the 35,000–65,000 range and they can be isolated retaining ligand-binding activity in a highly purified state.

Although C3b was a more effective inhibitor in dose-dependent inhibition experiments than C3, both C3 and C3b bound to the putative complement receptor isolated from rabbit macrophages. Solubilization could possibly affect the observed specificity of the receptor as previously discussed^{12,14}. That C3 does bind directly to intact cells has been documented previously for lymphocytes¹⁷, but has not been evaluated for macrophages. Based on these data there may be a parallel between the IgG-Fc γ receptor interaction and the interaction of C3 and C3b with their receptor. Although both C3 and IgG can bind to their respective receptors, they exhibit a higher net affinity of binding on modification: by fixation to antigen allowing multi-point attachment in the case of IgG, or by cleavage and removal of the C3a fragment in the case of C3. In addition, receptor-bearing effector cells may interact more efficiently with membrane-bound C3b molecules clustered about an antigenic site than C3 molecules in solution.

Previous attempts to purify the C3b receptor from solubilized cells have been unsuccessful^{1,18}. Recently, using affinity chromatography with C3-Sepharose, Fearon¹⁹ isolated a 205,000-MW protein from human erythrocytes which has been implicated as the C3b receptor²⁰, although ligand-binding activity of the purified material was not reported. Its relationship to the complement-binding macrophage glycoprotein described here remains to be defined.

We acknowledge the technical assistance of V. Krause and thank Drs C. Hammer and M. Frank for providing purified C3 for use in preliminary experiments.

Received 2 July 1980; accepted 2 February 1981.

1. Bianco, C. in *Biological Amplification Systems in Immunology* (eds Day, N. K. & Good, R. A.) 69–84 (Plenum Medical Book Co., New York, 1977).
2. Gelfand, M. C., Frank, M. M. & Green, I. J. *exp. Med.* **142**, 1029–1034 (1975).
3. Tack, B. F. & Prahl, J. W. *Biochemistry* **15**, 4513–4521 (1976).
4. Law, S. K., Lichtenberg, N. A. & Levine, R. P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7194–7198 (1980).
5. Bokisch, V. A., Muller-Eberhard, H. J. & Cochrane, C. G. *J. exp. Med.* **129**, 1027–1048 (1969).

6. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
7. Kulczycki, A. Jr, Krause, V., Chew Killion, C. & Atkinson, J. P. *J. Immun.* **124**, 2772–2779 (1980).
8. Cuatrecasas, P. *J. biol. Chem.* **245**, 3059–3065 (1970).
9. Kulczycki, A. Jr & Parker, C. W. *J. biol. Chem.* **254**, 3187–3193 (1979).
10. Kulczycki, A. Jr, Krause, V., Chew Killion, C. & Atkinson, J. P. *J. immun. Meth.* **37**, 133–138 (1980).
11. Pepys, M. B., Bell, A. J. & Rowe, I. R. *Scand. J. Immun.* **4**, Suppl. 3, 75–78 (1975).
12. Kulczycki, A. Jr, Hempstead, B. L., Hofmann, S. L., Wood, E. W. & Parker, C. W. *J. biol. Chem.* **254**, 3194–3200 (1979).
13. Schneider, R. J., Atkinson, J. P., Krause, V. & Kulczycki, A. Jr *J. Immun.* **126**, 735–740 (1981).
14. Hempstead, B. L., Parker, C. W. & Kulczycki, A. Jr *J. Immun.* **123**, 2283–2291 (1979).
15. Ishizaka, T., Ishizaka, K., Conrad, D. H. & Froese, A. *J. Allergy clin. Immun.* **61**, 320–330 (1978).
16. Isersky, C., Taurag, J. D., Poy, G. & Metzger, H. *J. Immun.* **121**, 549–558 (1978).
17. Theofilopoulos, A. N., Bokisch, V. A. & Dixon, F. J. *J. exp. Med.* **139**, 696–711 (1974).
18. Dierich, M. P. & Reisfeld, R. A. *J. Immun.* **114**, 1676–1682 (1975).
19. Fearon, D. T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5867–5871 (1979).
20. Fearon, D. T. *J. exp. Med.* **152**, 20–30 (1980).
21. Davis, B. J. *Ann. N. Y. Acad. Sci.* **121**, 404–427 (1964).

Cyclosporin A blocks receptors for HLA-DR antigens on T cells

Ronald Palacios* & Göran Möller

Department of Immunobiology, Karolinska Institute, Wallenberg Laboratory, Lilla Frescati, S-104 05, Stockholm, Sweden

Recent studies have shown that cyclosporin A, an undecapeptide extracted from two species of fungi (*Cylindrocarpum lucidum* and *Tricoderma polysporum*), has potent immunosuppressive effects with low myelocytotoxicity¹. In different experimental systems in both animals and humans, cyclosporin A administration decreased antibody production to T cell-dependent antigens, prolonged survival of skin, kidney and heart allografts^{1–3}, inhibited cytotoxic activity generated in the allogeneic mixed lymphocyte reaction⁴ and abrogated suppressor T-cell activity induced by concanavalin A (ref. 5). However, the mode of action of cyclosporin A and its role in establishing tolerance remains unclear. Continuous proliferation of T cells after stimulation with antigens or mitogens is maintained by growth factors, released by lectin- or antigen-activated T cells^{5,6}. One of these factors, interleukin 2 (IL-2), supports growth of activated but not of resting T cells^{5–6}, so that, to achieve a proliferative response a cell population must be 'activated' both to produce IL-2 and to become responsive to it. In the autologous mixed lymphocyte reaction (AMLR), purified human T lymphocytes proliferate when co-cultured with autologous B cells and/or macrophages^{7,8}. Recently, we found that antibodies against either the heavy non-polymorphic chain or the light polymorphic chain of HLA-DR antigens of the stimulator cells, prevented T cells from acquiring responsiveness to IL-2, inhibited the production of the growth factor and abrogated the T-cell proliferative response in AMLR and hapten-labelled autologous cell systems^{9,10}. These results suggested that the HLA-DR antigens rendered resting T cells sensitive to IL-2 and participated actively in the production of the growth factor. We have now studied the effect of cyclosporin A on the AMLR response, and found that the drug inhibited the proliferative response by blocking the receptors for HLA-DR antigens on T cells.

The addition of as little as 1 $\mu\text{g ml}^{-1}$ cyclosporin A to the AMLR cultures strongly inhibited the proliferative response when compared with control cultures. This inhibition occurred very early in the process of activation of the responding cells, for it was not seen if the drug was added 72 h after the initiation of the cultures (Fig. 1). Pretreatment of the T cells with 1 $\mu\text{g ml}^{-1}$ or more cyclosporin A abrogated proliferation in the AMLR, whereas treatment of the stimulator cells with up to 5 $\mu\text{g ml}^{-1}$ had no effect on the AMLR response. This indicates that cyclosporin A exerts its inhibitory effects by acting on the responding T cells rather than on the non-T stimulator cells, in

* To whom all correspondence should be sent.

agreement with previous studies which postulated a selective activity of cyclosporin A on T cells^{2,4}.

To define the mechanism by which cyclosporin A inhibits the AMLR response, we studied the sensitivity of the responding T cells to IL-2 in the presence or absence of the drug and its effect on the synthesis of IL-2 in the AMLR. The results show that T cells from AMLR cultures containing cyclosporin A did not respond to IL-2 stimulation, whereas T cells from cultures where no drug was added proliferated under the stimulus of IL-2 (Table 1). When previously activated T cells (by AMLR or by phytohaemagglutinin, PHA) were stimulated with different concentrations of IL-2 in the presence or absence of cyclosporin A inhibition did not occur, indicating that once T cells have responded to IL-2, they are no longer sensitive to the inhibitory effects of cyclosporin A (data not shown). These results, together with our previous findings that cyclosporin A-treated T cells are unable to absorb IL-2 whereas activated T cells are able to do so⁵, support the notion that the drug inhibits expression of IL-2 receptors on T cells^{5,11}. The production of IL-2 in the AMLR was also inhibited by the addition of cyclosporin A to the cultures (Table 2), suggesting a second mechanism by which this drug could inhibit T-cell activation in the AMLR.

It seems unlikely that cyclosporin A exerts its inhibitory effects by nonspecific toxic or lytic effects, for at the doses used in this study (1–5 $\mu\text{g ml}^{-1}$) the viability of the cells as determined by Trypan blue exclusion from AMLR cultures in the presence or absence of cyclosporin A were similar at the end of the culture period. Furthermore, the experiments carried out to determine

Table 2 Effect of cyclosporin A on the synthesis of IL-2 in the AMLR

Supernatant source	Dilutions of IL-2		
	1:2	1:4	1:8
T cells culture alone	1.1 $\pm$ 0.2*	1.0 $\pm$ 0.1	1.0 $\pm$ 0.2
AMLR	15.2 $\pm$ 2.4	10.6 $\pm$ 1.5	8.4 $\pm$ 0.9
AMLR + 1 $\mu\text{g ml}^{-1}$ cyclosporin A	2.9 $\pm$ 0.7	1.9 $\pm$ 0.2	1.8 $\pm$ 0.4
AMLR + 3 $\mu\text{g ml}^{-1}$ cyclosporin A	1.6 $\pm$ 0.5	1.8 $\pm$ 0.3	1.5 $\pm$ 0.2

IL-2 production. IL-2 was obtained from AMLR cultures as previously described⁵. Briefly, 3×10^6 T cells and 3×10^6 irradiated autologous non-T cells suspended in 1 ml of CCM in the presence or absence of 1 and 3 $\mu\text{g ml}^{-1}$ of cyclosporin A were incubated in 17×100 -mm round-bottom plastic tubes (Falconplastics) at 37 °C for 48 h. After the first 24 h of culture, the cells were washed three times in CCM, resuspended in CCM in the absence of cyclosporin A and incubated for the last 24 h of culture. After this, the tubes were centrifuged at 2,000 r.p.m. for 15 min, and the supernatants collected, filtered through 0.45 μm and stored at –20 °C until used. The experiments performed to find out whether cyclosporin A inhibits the proliferative response of previously activated T cells to IL-2 stimulation were established as follows: 10^4 T cells, activated in AMLR or by PHA and maintained in culture with conditioned media containing IL-2 for at least 3 weeks, were placed in flat-bottom microtitre plates. Three different dilutions of AMLR supernatants with proved IL-2 activity or IL-2 obtained after 48 h of PHA stimulation of peripheral blood MNC as described by Bonnard *et al.*¹⁶ was added with cyclosporin A (1 and 3 $\mu\text{g ml}^{-1}$) or medium to each well. Cultures in triplicate were incubated at 37 °C for 48 h and proliferation during the last 12 h was determined by the incorporation of ^3H -thymidine as described elsewhere^{5,7}. Supernatants from the following cultures: T cells cultured but not stimulated, AMLR = T cells and irradiated autologous non-T cells, AMLR + 1 μg or 3 $\mu\text{g ml}^{-1}$ cyclosporin A = T cells co-cultured with irradiated autologous non-T cells in the presence of cyclosporin A, were placed at three different final dilutions in each well of microtitre plates containing 10^4 T cells activated by PHA and maintained in culture with conditioned media plus IL-2 for 3 weeks. After 48 h of incubation, incorporation of ^3H -thymidine was determined (see Table 1 legend).

* c.p.m. $\times 10^{-3}$, mean $\pm$ s.d. of seven separate experiments.

the effects of this drug on the response to IL-2 stimulation of previously activated T cells revealed that cyclosporin A did not inhibit the proliferation of these cells after IL-2 stimulation. Moreover, the observation (Fig. 1) that T-cell proliferation was not inhibited when cyclosporin A was added 72 h after the initiation of AMLR cultures eliminates the possibility of nonspecific toxic or lytic effects on the responding cells. Finally, activation of the stimulator cells with Epstein–Barr virus was not affected by the presence of cyclosporin A in the cultures, and pretreatment of the stimulator cells with this drug did not modify their capacity to induce proliferation (data not shown).

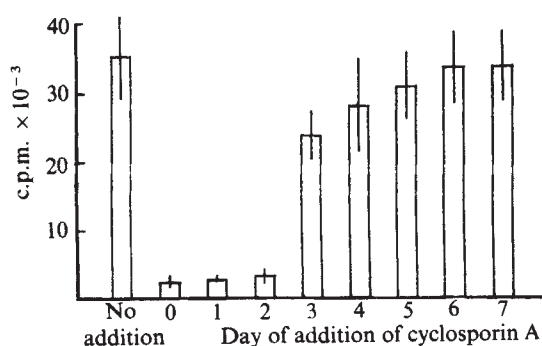


Fig. 1 Kinetic study of the inhibitory effect of cyclosporin A on the AMLR response. AMLR between 10^5 T cells and 10^5 irradiated autologous non-T cells were co-cultured in a final volume of 0.2 ml of complete culture medium (see Table 1 legend). Cyclosporin A (1 $\mu\text{g ml}^{-1}$) or medium was added to the cultures at different intervals after initiation of the AMLR cultures. Proliferation was determined by the ^3H -thymidine uptake during the last 12 h of the 7-day culture period.

Table 1 Response to IL-2 of T cells from AMLR cultures performed in the presence or absence of cyclosporin A

Cells	Dilutions of IL-2		
	1:2	1:4	1:8
T cells cultured alone	1.1 $\pm$ 0.1*	1.0 $\pm$ 0.3	0.9 $\pm$ 0.1
T AMLR	9.1 $\pm$ 1.3	6.5 $\pm$ 0.2	4.1 $\pm$ 0.5
T AMLR + 1 $\mu\text{g ml}^{-1}$ cyclosporin A	1.3 $\pm$ 0.4	0.8 $\pm$ 0.4	0.8 $\pm$ 0.2
T AMLR + 3 $\mu\text{g ml}^{-1}$ cyclosporin A	1.0 $\pm$ 0.3	0.7 $\pm$ 0.2	0.9 $\pm$ 0.1

Cell separation: Peripheral blood mononuclear cells (MNC) from healthy adult volunteers were obtained as previously described^{5,7}. T and non-T cells were isolated by rosetting MNC with sheep red blood cells (SRBC) as described elsewhere⁷. After osmotic lysis of the erythrocytes, the lymphocytes were washed and resuspended in RPMI 1640 medium (Gibco) supplemented with 10% of heat-inactivated human AB serum, 20 $\mu\text{g ml}^{-1}$ of gentamicin, glutamine and 1% of HEPES buffer solution (Gibco) which will be referred to as complete culture medium (CCM). Autologous mixed lymphocyte reaction cultures: AMLR cultures were performed as previously described⁷. Briefly, 10^5 T cells were cultured with 10^5 irradiated autologous non-T cells in round-bottom microtitre plates (Nunc) Delta, Nunc). The secondary AMLR cultures were established by co-culturing 10^5 T cells from each primary AMLR culture with 10^5 of the following irradiated (2,500 rad) autologous stimulator cells: non-T cells, T cells activated by PHA for 8 days (T-PHA), T cells from AMLR cultures performed in the presence or absence of cyclosporin A and peripheral blood MNC stimulated with PHA for 4 days (MNC-PHA). Cultures in triplicate were incubated at 37 °C in a humidified atmosphere of 5% CO_2 in air. 0.5 μCi of ^3H -thymidine (Radiochemical Centre) was added to each well 12 h before the end of a 7-day culture for primary AMLR cultures and a 3-day culture period for secondary cultures. ^3H -thymidine uptake of the cultures was determined as previously described^{5,7}. Cyclosporin A was tested for its ability to inhibit the proliferative response in AMLR by adding different doses of the drug (1–5 $\mu\text{g ml}^{-1}$) to the cultures either at the initiation of or at different intervals during culture. In no instance was the concentration of ethanol (in which cyclosporin A was dissolved) greater than 0.25% whether in the cultures containing this drug or in medium (control). IL-2 obtained from PHA-stimulated MNC was added to 10^2 T cells of the following groups: T cells cultured but not stimulated, T AMLR = T cells from AMLR cultures, T AMLR + 1 μg or 3 μg cyclosporin A = T cells from AMLR cultures treated with cyclosporin A. Proliferation was determined by the incorporation of ^3H -thymidine during the last 12 h of the 2-day culture period.

* c.p.m. $\times 10^{-3}$, mean $\pm$ s.d. of seven independent experiments.

Table 3 Response to various autologous stimulator cells of T cells from primary AMLR cultures performed in the presence or absence of cyclosporin A

Primary culture	Cyclosporin A addition ($\mu\text{g ml}^{-1}$)	Secondary AMLR culture stimulator cells				
		Non-T	T-PHA	T-AMLR	MNC-PHA	T-(AMLR + CYA)
T cells culture alone	—	1.3 $\pm$ 0.1*	1.8 $\pm$ 0.2	1.9 $\pm$ 0.1	2.4 $\pm$ 0.3	0.7 $\pm$ 0.1
	1	1.0 $\pm$ 0.1	0.7 $\pm$ 0.1	0.6 $\pm$ 0.2	0.9 $\pm$ 0.1	0.5 $\pm$ 0.1
T AMLR	—	33.4 $\pm$ 2.8	29.3 $\pm$ 1.7	20.6 $\pm$ 2.5	31.5 $\pm$ 1.9	2.9 $\pm$ 0.7
	1	6.3 $\pm$ 1.7	5.0 $\pm$ 1.8	4.8 $\pm$ 2.0	7.6 $\pm$ 2.3	1.1 $\pm$ 0.2
	3	4.2 $\pm$ 1.5	3.5 $\pm$ 0.6	3.3 $\pm$ 1.1	4.1 $\pm$ 1.6	1.4 $\pm$ 0.5
T AMLR + 1 $\mu\text{g ml}^{-1}$ cyclosporin A	—	1.9 $\pm$ 0.1	1.3 $\pm$ 0.1	1.5 $\pm$ 0.3	4.2 $\pm$ 1.0	0.6 $\pm$ 0.1
T AMLR + 3 $\mu\text{g ml}^{-1}$ cyclosporin A	—	1.1 $\pm$ 0.3	1.0 $\pm$ 0.1	1.4 $\pm$ 0.4	4.1 $\pm$ 0.7	0.9 $\pm$ 0.3

Cyclosporin A (1 or 3 $\mu\text{g ml}^{-1}$) or medium was added to each well at the initiation of the secondary AMLR culture. T cells cultured alone or with irradiated autologous non-T cells in the presence (T-AMLR + 1 or 3 $\mu\text{g ml}^{-1}$ cyclosporin A) or absence (T-AMLR) of cyclosporin A for 10 days (primary AMLR culture) were recovered by rosetting with SRBC and restimulated with the following autologous cells: Non-T cells; T-PHA = 8-day PHA-stimulated T cells; T-AMLR = T cells from 7-day cultures; MNC-PHA = 4-day PHA-stimulated peripheral blood MNC; T-AMLR = T cells from 7-day AMLR culture done in the presence of 1 $\mu\text{g ml}^{-1}$ cyclosporin A. The cell mixture was incubated at 37 °C for 3 days, and the ^3H -thymidine incorporation during the last 12 h of the culture period determined (see Table 1 legend).

* Mean $\pm$ s.d. of three independent experiments expressed as c.p.m. $\times 10^{-3}$.

We next studied the effects of cyclosporin A on the secondary AMLR using different kinds of irradiated autologous cells as stimulators. T cells activated in primary AMLR cultures or by PHA possess DR antigens on their surface^{9,12,13} and can function as stimulator cells in the AMLR⁹. Table 3 shows that the addition of cyclosporin A at the start of the secondary AMLR cultures inhibited the proliferation of the responding T cells. Interestingly, T cells from primary AMLR cultures performed in the presence of cyclosporin A were unable to function as stimulators in secondary AMLR cultures, whereas T cells from primary cultures without cyclosporin A functioned well as stimulator cells. Furthermore, the former cells were not killed by treatment with a monoclonal anti-DR antibody (NE011, NEN) plus complement whereas 50–60% of the latter cells were lysed by similar treatment. This indicates that T cells exposed to cyclosporin A lose their capacity to bind DR antigens, a property of T cells which has been demonstrated by Elliot *et al.*¹⁴

We also found that T cells from primary cultures containing cyclosporin A exhibited a proliferative activity similar to that shown by the population of cultured, not stimulated, T cells, whereas non-cyclosporin A-treated T cells responded with a typical pattern of secondary responses (higher proliferation at an earlier time). These findings suggest that cyclosporin A-treated T cells were not activated in the first culture, as they proliferate in secondary cultures to a similar extent to T cells cultured but not activated.

Thus, cyclosporin A inhibits the AMLR response by acting selectively on the responding T cells before they become responsive to IL-2, after which the drug can no longer inhibit T-cell proliferation and/or growth. We therefore conclude that cyclosporin A exerts its inhibitory effects by interfering with the process which normally enables resting T cells to become sensitive to IL-2, a step which is induced by signals from HLA-DR antigens of the stimulator cells^{9,10}. As cyclosporin A has no effect on stimulator cells, it probably blocks the receptors for the DR molecules on T cells and consequently inhibits the expression of IL-2 receptors on these cells. This notion is supported by several pieces of evidence. First, cyclosporin A-treated T cells are unresponsive and unable to absorb IL-2 (ref. 5 and the present study). Second, T cells from primary AMLR cultures performed in the presence of cyclosporin A do not possess DR antigens and are unable to function as stimulator cells in the AMLR. Third, cyclosporin A does not affect other receptors present on T cells, as it did not reduce the number of T cells forming rosettes with sheep red blood cells or autologous erythrocytes (data not shown).

We have recently found that HLA-DR antigens participate in the process of IL-2 production in either the AMLR or allogeneic mixed lymphocyte reaction (MLR) by rendering IL-2-producer T cells sensitive to interleukin 1 (formerly lymphocyte activating factor), which induces these cells to secrete the growth factor. The addition of cyclosporin A to AMLR and MLR cultures

causes IL-2-producer T cells to become unresponsive to interleukin 1 and abrogates the synthesis of IL-2¹⁵.

This work was supported by grants from the Swedish Cancer Society and the Swedish MRC.

Received 13 October 1980; accepted 11 February 1981.

- Borel, J. F., Feuer, C., Gubler, H. U. & Stahelin, H. Ag. *Actions* **6**, 468–475 (1976).
- Borel, J. F. *Immunology* **31**, 631–641 (1976).
- White, D. J. G., Calne, R. Y. & Plumb, A. *Transplant Proc.* **11**, 855–859 (1979).
- Horsburgh, T., Wood, P. & Brent, L. *Nature* **286**, 609–611 (1980).
- Palacios, R. & Möller, G. *J. exp. Med.* **153**, (1981).
- Smith, K. A. *Immun. Rev.* **51**, 338–357 (1980).
- Palacios, R., Llorente, L., Alarcón-Segovia, D., Ruiz-Argüelles, A. & Diaz-Jouanen, E. *J. clin. Invest.* **65**, 1527–1530 (1980).
- Opelz, G., Takasugi, M. & Terasaki, P. I. *J. exp. Med.* **142**, 1327–1333 (1975).
- Palacios, R. *et al.* (submitted).
- Palacios, R., Möller, G., Claesson, L. & Peterson, P. A. (submitted).
- Larsson, E. L. *J. Immun.* **124**, 2828–2833 (1980).
- Elliot, B. E., Nagy, Z., Nabholz, M. & Pernis, B. *Eur. J. Immun.* **7**, 287–291 (1977).
- Engleman, E. G., Benike, C. J. & Charron, D. J. *J. exp. Med.* **152**, Suppl., 114–126 (1980).
- Elliot, B. E., Nagy, Z. A., Ben-Neriah, Y. & Givol, D. *Nature* **285**, 496–498 (1980).
- Palacios, R. (submitted).
- Bonnard, G. D., Yasaka, K. & Maca, R. D. *Cell. Immun.* **51**, 390–401 (1980).

Role of RNA transcripts in replication incompatibility and copy number control in antibiotic resistance plasmid derivatives

Jonathan Rosen*, Thomas Ryder, Hisako Ohtsubo & Eiichi Ohtsubo

Department of Microbiology, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

The genes required for autonomous replication and incompatibility in the antibiotic resistance plasmids R100 and R1 have been located within a 2.5-kilobase region of the 90-kilobase genome, within which the incompatibility gene occupies a 1.3-kilobase region excluding the replication origin. We now report that three RNA species are synthesized *in vitro* from the 2.5-kilobase region, which R100 and R1 have in common. One, a long RNA molecule which is transcribed in the direction of DNA replication, probably acts as a messenger or a protein required for plasmid replication. The second RNA species, only 91 nucleotides long, is transcribed in the opposite direction, from a region of the DNA entirely contained within the first and known to specify incompatibility and copy control functions. The third RNA species, 150 bases long, is transcribed from a region including the replication origin¹; it may be a primer of DNA synthesis or, in conjunction with the second of the three RNA species, an influence in the control of replication.

The various components involved in replication control were studied using the small multicopy derivatives of R100 and R1, pSM1 and pTR1, respectively (Fig. 1). The nucleotide sequence of the entire region for replication of pSM1 (ref. 2) and pTR1 (T.R. *et al.*, in preparation) was previously determined. *In vitro* transcription of whole pSM1 and pTR1 and the *Pst*I fragment known to be responsible for incompatibility (see Fig. 1) produced a small densely labelled RNA species common to pSM1 and pTR1 when synthesized in the presence of [γ - 32 P]ATP or [α - 32 P]UTP. We call this transcript RNAI, and sequence analysis (see Fig. 2) reveals that the 5' end occurs at nucleotide 563 and that it generally terminates at nucleotide 473 of the previously determined pSM1 replication region sequence (Fig. 3). The RNA can form two large stable secondary structures, as shown in Fig. 4. Interestingly the two differences in sequence between pTR1 and pSM1 in this region occur at the top of the large secondary structure shown in Fig. 4, while 13 of the first 14 nucleotides of RNAI are complementary to the sequence between 2,367 and 2,380 of the replication origin region (Figs 3, 4).

In vitro transcription of pSM1 or pTR1 in the presence of [γ - 32 P]GTP or [α - 32 P]UTP produces a much larger RNA species common to pSM1 and pTR1, which we call RNAII. The 5' end of this transcript occurs at nucleotide 392 but the location of the 3' end is unknown.

Both RNAI and RNAII are preceded by regions of reasonable homology to the prototypical RNA polymerase recognition region known as the -35 region (tgTTGACA-ttt) and Pribnow sequence (TATPuATPu)³ as shown in Fig. 3. Interestingly, the nucleotide sequence of the -35 region of the promoter for RNAII in pSM1 is different from that in pTR1, although the Pribnow sequence is the same.

When the *Pst*I fragment which carries the origin of replication of pSM1 or pTR1 is transcribed *in vitro* in the presence of

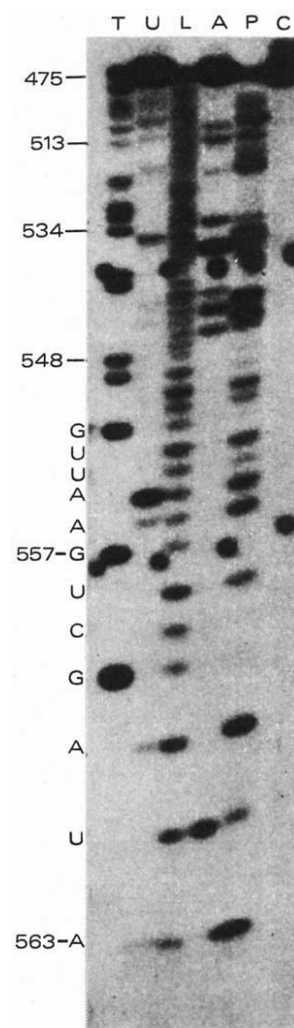


Fig. 2 Autoradiogram of a sequencing gel of the 5' end of RNAI. The position of selected G residues on the sequence shown in Fig. 3 are marked. Sequencing was performed by enzymatic cleavage as described by Donnis-Keller *et al.*¹⁵ and Simoncsits *et al.*¹⁶. The lanes of the gel are as follows: C, control; P, *PhyI*; A, pancreatic RNase A; L, alkaline digest; U, RNase U₂; T, RNase T₁. RNA was synthesized *in vitro*, basically as described by Majors¹⁷ in a mixture containing 0.2 pmol DNA (circular plasmids or linear restriction fragments) 30 mM Tris (pH 8.0), 2 mM MgCl₂, 0.1 M KCl, 0.1 mM EDTA, 0.1 mM dithiothreitol and 0.5 mg ml⁻¹ bovine serum albumin. RNA polymerase (Boehringer Mannheim) was added in a molar ratio of 10:1 enzyme to DNA and the mixture was incubated for 10 min at 37 °C. Nucleotides were added to a final concentration of 20 μ M except for the labelled nucleotide, which was added to 2 μ M. The mixture was then transferred to a tube containing 40 μ Ci [α - 32 P]UTP or 0.2 mCi dried down [γ - 32 P]ATP¹⁸ or [γ - 32 P]GTP (>1,000 Ci mmol⁻¹, ICN) and incubated at 37 °C for 30 min. The transcripts were purified by gel electrophoresis as described by Donnis-Keller *et al.*¹⁵.

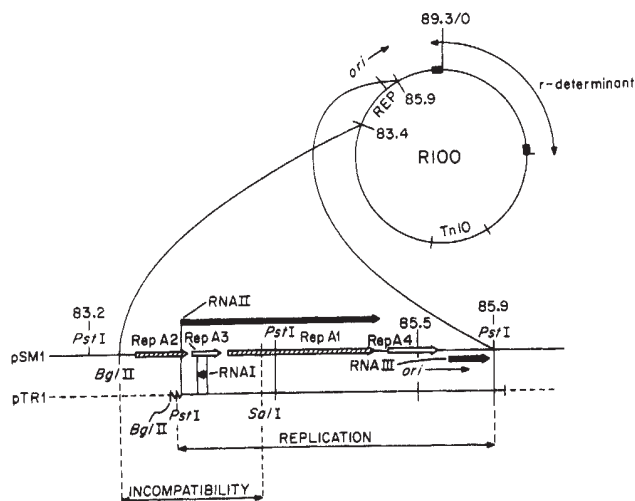


Fig. 1 A circular representation of some genetic and physical properties of the plasmid R100. The region required for replication (REP) includes the origin of replication (*ori*) from which replication proceeds unidirectionally in the direction shown. The origin region was determined by electron microscopic analysis of replicating molecules¹⁴ and the sequence was determined¹. The figure also shows a linear representation of the replication regions of the R100 derivative pSM1 and the R1 derivative pTR1 which was constructed from Rsc13 such that only the replication region and the gene conferring ampicillin resistance remain (T.R., unpublished results). The regions required for replication and incompatibility are shown at the bottom of the figure. The open arrows marked RepA3 and RepA4 indicate open reading frames predicted by nucleotide sequence analysis². The cross-hatched arrows show the position of reading frames whose polypeptide products have been confirmed². The heavy arrows marked RNAI, RNAII and RNAIII are the transcripts described in the text.

[γ - 32 P]ATP, a third transcript called RNAIII is found. The 5' end of this transcript occurs at nucleotide 2,316, and the region upstream of this site displays moderate homology to prototypical promoter sites (Fig. 3). The transcript has two termination sites when this *Pst*I fragment is used as a template, the first being simply the end of the *Pst*I fragment, and the second at approximately nucleotide 2,465, about seven bases upstream of the *Pst*I cutting site. Label is preferentially incorporated in the shorter transcript with a ratio of 3.7:1. The sequence in this region, TAATCAATAT, is remarkably similar to that at the termination site of λ t_{R1}-CAATCAA⁴-and tRNA^{Tyr}-CAATCAATAT⁵-both of which are ρ -factor-dependent transcriptional termination sites.

* Present address: Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA.

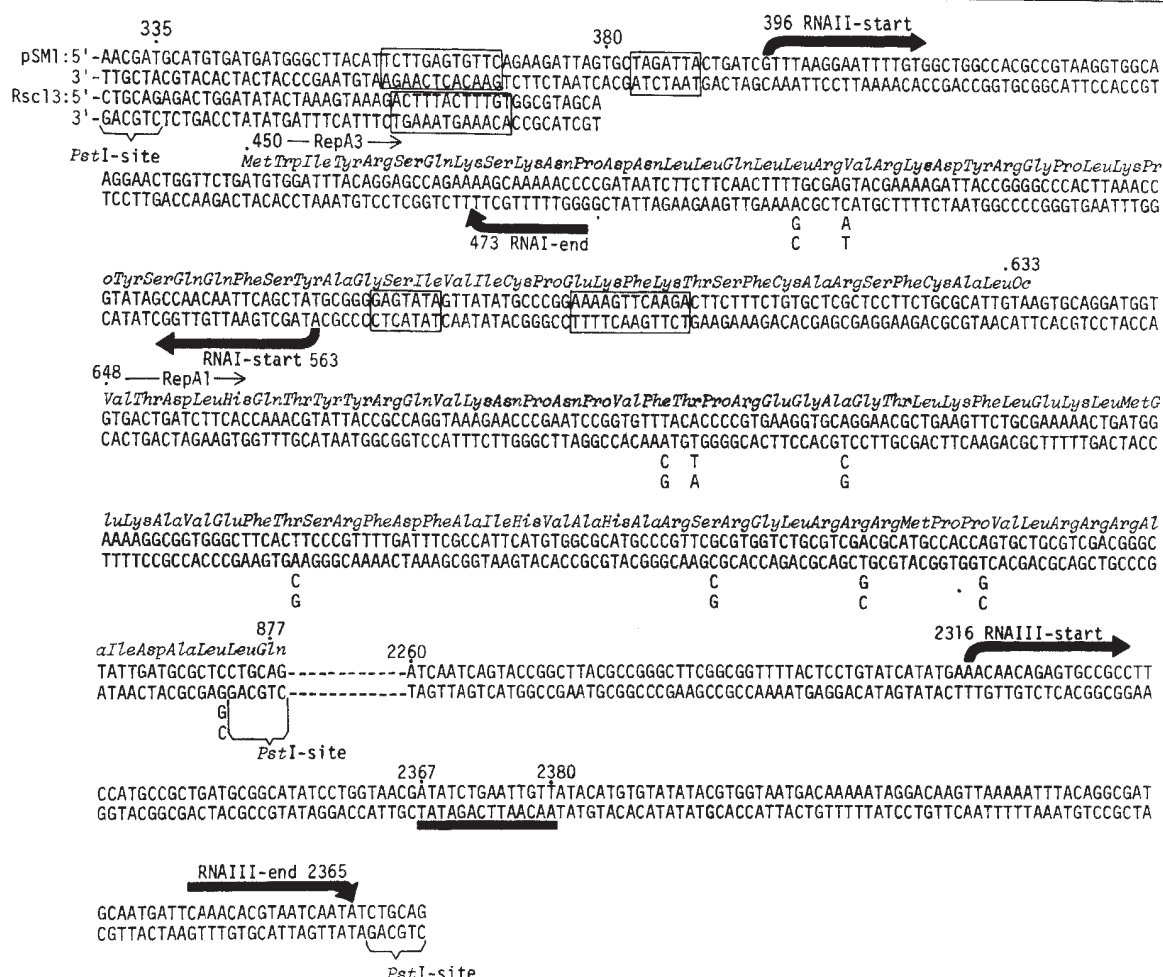


Fig. 3 The location of RNAI and RNAII within a *Pst*I fragment of pTR1. The entire sequence of the plasmid pSM1 within this fragment is shown and the differences between pSM1 and pTR1 are shown below the pSM1 sequence. The heavy lines above and below the sequence show the location of RNAI and RNAII at the nucleotide sequence level. The boxed regions proximal to each transcript delineate the Pribnow and -35 recognition sequences. The coding regions for RepA1 and RepA3 as well as the amino acid residues encoded are indicated on the sequence. A portion of the sequence of the origin region of pSM1 (2,260–2,480) including RNAIII and the sequence complementary to the 5' end of RNAI is also shown.

RNAIII has only been found when the linear *Pst*I fragment is transcribed, and its absence when superhelical pSM1 or pTR1 DNA is used as template may be due to nonspecific termination in the absence of ρ -factor yielding a heterogeneous range of transcripts. Most interestingly, the region of complementarity between RNAI and the origin region occurs within this transcript (Fig. 3). As will be discussed below, the interaction between RNAI and the superhelical template may affect the transcription of RNAIII.

RNAII is almost certainly the messenger RNA required to translate a 33,000-molecular weight (MW) protein called RepA1 which is assumed to be necessary for replication initiation (see Fig. 1 and ref. 2). As RNAII spans the coding region of a second, much smaller hypothetical 7,000-MW protein called RepA3, it may also direct the synthesis of that protein. However, RepA3 has not been seen *in vivo* or *in vitro*, so it is unclear whether this hypothetical protein is translated.

An interesting hypothesis is that RNAI may be partly responsible for copy number control and incompatibility, because it is synthesized in the region responsible for these functions. Figures 3 and 4 show that the two sequence differences between pSM1 and pTR1 in this region occur at the top of the proposed secondary structure for RNAI. Comparison of these sequences with the wild-type R1 DNA sequence recently obtained by Stougaard *et al.*⁶ reveals a single base-pair change between wild type and pSM1 at position 505 and

between wild type and pTR1 at position 510; these two differences between the pSM1 and pTR1 DNA sequences are probably responsible for their mutant copy number phenotypes, possibly altering the copy number and incompatibility properties of these plasmids by altering the affinity that RNAI has to a protein also involved in replication control.

RNAIII is likely to act either as the actual primer for the initiation of the plasmid DNA synthesis, or as a transcriptional activation signal which we suggest alters the conformation of a secondary structure at the replication origin upstream of this RNA transcript much as proposed by Hobom *et al.* for λ replication⁷. A large secondary structure has in fact been found upstream of RNAIII in pSM1 (ref. 1). As described above, 13 of the first 14 nucleotides of RNAI are complementary to the region between 2,367 and 2,380 of the replication origin region spanned by RNAIII (Figs 3, 4). Assuming that RNAI is involved in the incompatibility reaction and acts directly at the origin, RNAI may inhibit the initiation of DNA synthesis by binding to the complementary sequence at the origin region, probably with the aid of a protein(s).

In our *in vitro* system, RNAI and RNAII are transcribed simultaneously from the same region of the plasmid but in opposite directions. *In vivo*, the production of RNAI itself may be controlled at the transcriptional level by an attenuator-like mechanism^{8,9} affecting both copy number and incompatibility. If a ribosome succeeds in binding to a nascent RNAI before the

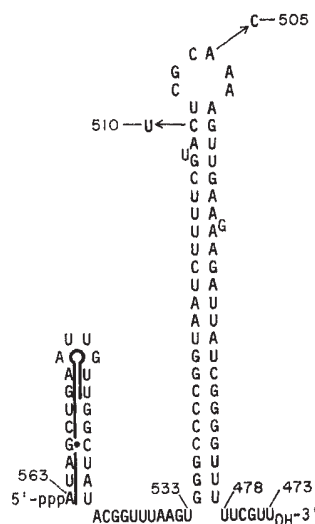


Fig. 4 Possible secondary structure of RNAI and homology to the origin of pSM1. The two arrows at the top of the loop indicate the differences between the sequence of RNAI and pSM1 and pTR1 at positions 505 (A → C) and 510 (C → U). The line within the first small hairpin loop indicates the region complementary to the origin region between nucleotides 2,367 and 2,380 of the pSM1 replication region sequence (see Fig. 3 and ref. 1).

RNA polymerase reaches the termination point the secondary structure would be disrupted and, analogously to the pattern of transcription in attenuation systems^{8,9}, transcription might proceed beyond the termination point reported here. It is difficult to predict if RNAI is actually translated. A small open reading frame begins at the start codon GUG (535–533) just before the large stem structure seen in Fig. 4 and continues until the termination codon UAG (384–382) (see Fig. 3). This reading frame is preceded by the sequence GGU (542–540), which may serve as a ribosome binding site¹⁰. The protein which could be translated from the extended RNA would have a MW of ~5,300, and as this is the region encoding copy number control and incompatibility functions, both of which are probably controlled by a repressor from this region, the 5,300-MW polypeptide is a possible candidate. However, because the base-pair change between pSM1 and R1 would not cause an amino acid change if the extended RNAI were translated, this small polypeptide would be identical in the wild-type and mutant plasmids. The two modes of transcription of RNAI may, however, correspond to the repressed and derepressed states of replication. Thus, the rate of plasmid replication could be tied, deterministically, to other aspects of cellular metabolism by the average rate of initiation for RNAI translation. This larger RNA species may no longer be effective either because of steric problems in binding at the origin or in forming a duplex with nascent RNAII or because of improper or altered processing of the molecule.

Recent data obtained in the ColE1 plasmid system shows a very similar pattern of transcription in the region required for replication to that seen in pSM1 and pTR1 (ref. 11), although the ColE1 plasmid does not require any of its own proteins for replication and has no transcript corresponding to RNAIII. A small RNA which may be analogous to RNAI is produced in the direction opposite to the direction of replication and ~450 bases upstream from the replication origin¹². A small insertion mutation that maps in this RNA increases the copy number of the plasmid, indicating that this RNA is an important control element¹³. Furthermore, it is transcribed in the direction opposite that of a second, larger transcript required *in vitro* as a primer for DNA synthesis¹¹. We do not know whether RNAII in our system can be used as both an mRNA and a primer for R-factor replication. It is reasonable to propose that RNAI, by interacting at the origin region, is involved in copy number control,

based on its location in the replication region, the changes found in pSM1 and pTR1 that appear to cause the increased copy number phenotype and by partial analogy to the ColE1 system. Further genetic and biochemical analyses to determine the nature of this control system are under way.

We thank Chiyoko Machida for sharing her expertise in the use of various RNases, Bert Semler and Naomi and Akiko Kitimura for help in the initial stages of RNA sequencing, Yvonne Tong and Linda Hollman for typing the manuscript, Joyce Schirmer for the illustrations and Jeff Demian for the photographs. This work was supported by the USPHS with research grants to E.O. (GM22007) and H.O. (GM26779) and by partial support to T.R. under training grant (CA-09176).

Note added in proof: In recent attempts to shorten the replication region we have been unable to delete the region coding for RNAI and RNAII but the region coding for RNAIII can be eliminated. Thus RNAIII is unlikely to be a primer for DNA synthesis but the possibility that its normal function includes some other aspect of replication control is not excluded.

Received 23 September 1980; accepted 28 January 1981.

- Rosen J., Ohtsubo, H. & Ohtsubo, E. *Molec. gen. Genet.* **171**, 287–293 (1979).
- Rosen, J., Ryder, T., Inokuchi, H., Ohtsubo, H. & Ohtsubo, E. *Molec. gen. Genet.* **179**, 527–537 (1980).
- Rosenberg, M. & Court, D. A. *Rev. Genet.* **13**, 319–353 (1979).
- Rosenberg, M., Court, D., Shimatake, H., Brady, C. & Wulff, D. L. *Nature* **272**, 414–423 (1978).
- Küpper, H., Sekiya, T., Rosenberg, M., Egan, J. & Landy, A. *Nature* **272**, 423–428 (1978).
- Stougaard, P., Molin, S., Nordström, K. & Hansen, F. *Molec. gen. Genet.* (in the press).
- Hobom, G. et al. *Cold Spring Harb. Symp. quant. Biol.* **43**, 165–178 (1978).
- Lee, F. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4365–4369 (1977).
- Keller, E. B. & Calvo, J. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6186–6190 (1979).
- Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342–1346 (1974).
- Itoh, T. & Tomizawa, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2450–2454 (1980).
- Levine, A. D. & Rupp, W. D. in *Microbiology 1978* (ed. Schlessinger, D.) (American Society for Microbiology, Washington DC, 1978).
- Conrad, S. E. & Campbell, J. L. *Cell* **18**, 61–71 (1979).
- Ohtsubo, E., Feingold, J., Ohtsubo, H., Mickel, S. & Bauer, W. *Plasmid* **1**, 8–18 (1977).
- Donnis-Keller, H., Maxam, A. & Gilbert, W. *Nucleic Acids Res.* **4**, 2527–2538 (1977).
- Simoncsits, A., Brownlee, G. G., Brown, R. S., Robin, J. R. & Guillely, H. *Nature* **269**, 833–836 (1977).
- Majors, J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4394–4398 (1975).
- Johnson, R. A. & Walseth, T. F. *Adv. Cyclic Nucleotide Res.* **10**, 135–167 (1979).

An inducible DNA replication–cell division coupling mechanism in *E. coli*

Olivier Huisman & Richard D'Ari

Institut de Recherche en Biologie Moléculaire, 2, Place Jussieu (Tour 43), 75221 Paris Cedex 05, France

Cell division is a tightly regulated periodic process. In steady-state cultures of Enterobacteriaceae, division takes place at a well defined cell mass¹ and is strictly coordinated with DNA replication². In wild-type *Escherichia coli* the formation of cells lacking DNA is very rare³, and interruptions of DNA replication arrest cell division^{4–6}. The molecular bases of this replication–division coupling have been elusive but several models have been proposed. It has been suggested, for example, that the termination of a round of DNA replication may trigger a key event required for cell division^{4–6}. A quite different model postulates the existence of a division inhibitor which prevents untimely division and whose synthesis is induced to high levels when DNA replication is perturbed⁹. The work reported here establishes the existence of the latter type of replication–division coupling in *E. coli*, and shows that the *sfiA* gene product is an inducible component of this division inhibition mechanism which is synthesized at high levels after perturbations of DNA replication.

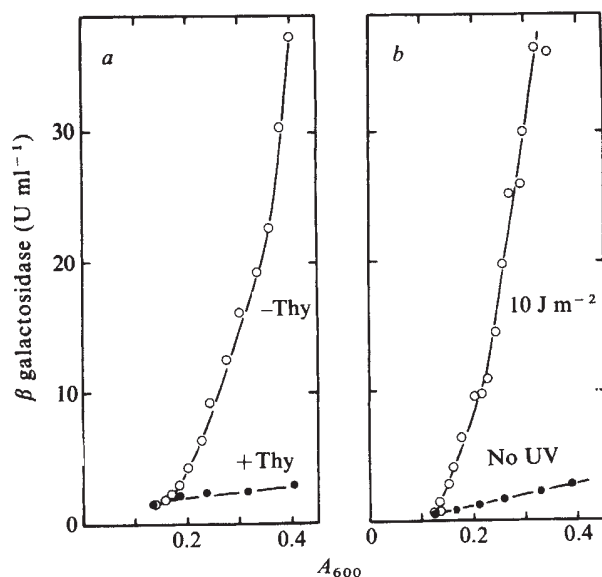


Fig. 1 Inducibility of *sfiA* gene expression by thymine starvation and UV-ray irradiation. Bacteria were grown at 37 °C in M63 minimal salts medium³⁵ supplemented with glucose, casamino acids and growth requirements. At an absorbance *A* of ~0.4, the cultures were adjusted to *A* 0.1, a part of each was induced (a, centrifuged and washed twice, then resuspended in medium without thymine; b, irradiated with 10 J m⁻² UV ray) and incubation was continued at 37 °C. Samples were withdrawn every 5 (induced cultures) or every 15 (uninduced cultures) min for absorbance measurements and β -galactosidase assays (carried out as in ref. 35). Enzyme units were calculated as described in ref. 37. Strain construction is described in Table 1 legend.

Much evidence points to the induction by DNA damaging treatments of a number of biological phenomena, collectively termed the SOS response (reviewed in refs 10–12). These phenomena include the appearance of new DNA repair and mutagenic activities, massive synthesis of the *recA* protein, lytic development of certain prophages in lysogenic strains and division inhibition (filamentous growth) in non-lysogens. The *recA* and *lexA* functions regulate the expression of all these phenomena. Mutant alleles of either locus can entirely prevent induction of the SOS response (*recA*⁻, *lexA*⁻) or cause its gratuitous expression at 42 °C (*tif* at the *recA* locus, *tsl* at the *lexA* locus)^{10–12}.

Incubation of non-lysogenic *tif* or *tsl* strains at 42 °C arrests cell division, resulting in the formation of long, multinucleate filaments. This filamentous growth is completely suppressed by *sfiA* mutations, whereas other aspects of the SOS response are unaffected^{13–15}. The *sfiA* locus thus defines an element directly involved in the division inhibition process.

One model of replication–division coupling predicts the existence of an inducible division inhibitor. We thus sought to determine whether synthesis of the *sfiA* product is induced when DNA replication is perturbed.

To obtain accurate quantitative measurements of the rates of *sfiA* gene expression in various conditions we chose the powerful method of operon fusion, using the phage Mu d(Ap, *lac*)¹⁶ to link the structural gene for β -galactosidase (*lacZ*) to the promoter of the *sfiA* operon. Near one extremity of the phage genome are the *lacZ* and *lacY* genes, lacking a promoter. The phage genome is so constructed that, when inserted in a host operon in the proper orientation, the *lac* genes of the prophage can be expressed from the adjacent bacterial promoter: their level of expression then reflects the level of expression of the bacterial operon in which the prophage is inserted.

Previous studies have shown that *sfi*-dependent filamentation occurs early during thymine starvation¹⁷. The effect of thymine starvation on *sfiA* expression was investigated in a *thyA sfiA::lacZ* fusion strain. The differential rate of β -galactosidase synthesis in these conditions quickly increases 20–40-fold (Fig. 1a, Table 1) reaching ~0.3% of total protein synthesis. Thus the *sfi*-dependent filamentation normally observed during thymine starvation is correlated with a large increase in the rate of *sfiA* gene expression. A UV-ray dose of 10 J m⁻² similarly causes a rapid 20–40-fold increase in the differential rate of β -galactosidase synthesis (Fig. 1b, Table 1). Nalidixic acid, a specific inhibitor of DNA gyrase, and methyl methanesulphonate, a DNA alkylating agent, also induce high rates of *sfiA* gene expression (Table 1). A Mu d(Ap, *lac*) prophage inserted in the *gal* operon does not lead to UV-ray inducible β -galactosidase synthesis (Table 1), showing that the inducibility observed in *sfiA::lacZ* strains depends on the *sfiA* location of the prophage. We conclude that perturbations of DNA replication induce a high level of expression of the *sfiA* gene.

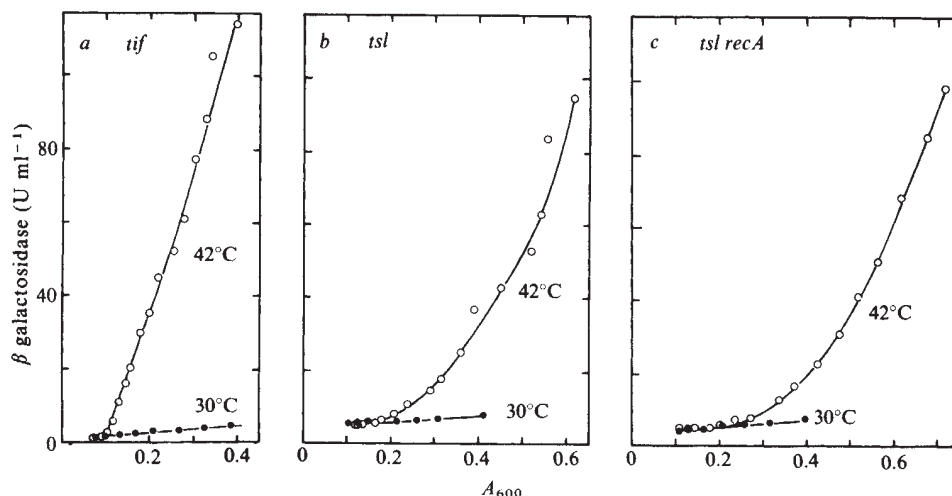
We next studied the effect on *sfiA* gene expression of mutations known to suppress or enhance *sfi*-dependent filamentation. *recA*⁻ and *lexA*⁻ mutations completely suppress the SOS

Table 1 Induction of *sfiA* gene expression

Strain	Inducing treatment	β -galactosidase specific activity (U per mg dry weight)		
		Un-induced	Induced 60 min	Dose
<i>sfiA::lacZ thyA</i>	Thymine starvation	50	410	
<i>sfiA::lacZ</i>	UV-ray irradiation	30	140	1 J m ⁻²
			440	10 J m ⁻²
<i>sfiA::lacZ</i>	Nalidixic acid	30	110	2 μ g ml ⁻¹
			710	40 μ g ml ⁻¹
<i>sfiA::lacZ</i>	Methyl methanesulphonate	30	360	0.025%
<i>sfiA::lacZ</i>	Nitrofurantoin	30	90	2 μ g ml ⁻¹
<i>sfiA::lacZ recA</i>	UV-ray irradiation	10	10	1 J m ⁻²
			10	10 J m ⁻²
<i>sfiA::lacZ lexA</i>	UV-ray irradiation	10	10	1 J m ⁻²
			10	10 J m ⁻²
<i>sfiA::lacZ tif</i>	42 °C	80	700	
<i>sfiA::lacZ tsl</i>	42 °C	200	450	
<i>sfiA::lacZ tsl recA</i>	42 °C	200	400	
<i>sfiA::lacZ lon</i>	UV-ray irradiation	20	400	10 J m ⁻²
<i>gal::lacZ</i>	Galactose	10	50	0.4%
	UV-ray irradiation	10	10	10 J m ⁻²
<i>gal::lacZ recA</i>	Galactose + UV rays	10	25	0.4%, 10 J m ⁻²
<i>gal::lacZ lexA</i>	Galactose + UV rays	10	30	0.4%, 10 J m ⁻²
<i>gal::lacZ tif</i>	42 °C	10	10	

Bacteria were grown at 37 °C in M63 salts medium³⁵ supplemented with glucose, casamino acids and growth requirements (except for the experiments with *tif* and *tsl* strains, grown at 30 °C) to a value of *A*₆₀₀ of ~0.4, then diluted to *A* 0.1 (the thymine-starved culture was first centrifuged and washed twice to remove exogenous thymine). One part of the culture was induced (by removal of thymine, UV-ray irradiation, temperature shift or addition of the appropriate factors), the cultures were further incubated for 60 min and β -galactosidase specific activities measured³⁵. Adenine (100 μ g ml⁻¹) was added to the *tif* cultures at the time of the temperature shift. To isolate Mu d(Ap, *lac*) insertions at the *sfiA* locus, we selected for a Sfi⁻ phenotype^{13,25,26} in a *lon* strain²⁴, taking advantage of the hypersensitivity of the strain to low concentrations of nitrofurantoin³⁶. Lysogenic Sfi⁻ clones were selected from a Mu d(Ap, *lac*)-infected *lon* culture on plates containing ampicillin (50 μ g ml⁻¹) and nitrofurantoin (2 μ g ml⁻¹). These were purified, checked for UV-ray resistance (to confirm the Sfi⁻ character) and analysed genetically. All Sfi⁻ clones carried a *sfiA* mutation, about 50% co-transducible with *pyrD* by P1^{13,15,25,26}. For most isolates the introduction by transduction of the Sfi⁺ character was invariably accompanied by loss of the Mu d(Ap, *lac*) prophage; the strains thus harboured a single Mu d(Ap, *lac*) prophage, located at the *sfiA* locus and conferring a Sfi⁻ phenotype. Strains of this type having a significant basal level of β -galactosidase were considered to have the *lac* genes fused to the *sfiA* promoter. One such fusion was transduced with P1 *vir* (selection Ap^R) into a strain carrying a chromosomal *lac* deletion (to prevent extraneous synthesis of β -galactosidase) and a Mu c⁺ prophage inserted in the *trp* operon (to prevent thermoinduction of the Mu d(Ap, *lac*) prophage, whose repressor is thermolabile). One Ap^R transductant, verified to have a Sfi⁻ phenotype, was used for this study; all other *sfiA::lacZ* fusion strains were derived from this transductant. The *gal::lacZ* strains were constructed in a similar way using a strain carrying Mu d(Ap, *lac*) inserted in the *galT* gene as donor in the transduction to Ap^R.

Fig. 2 Regulation of *sfiA* gene expression in *tif(recA)* and *tsl(lexA)* mutants. Bacteria were grown at 30 °C (a, in M63 minimal salts medium³⁵ supplemented with glucose, casaminoacids, growth requirements and—adenine; b and c, in LB broth). At A 0.4, the cultures were diluted to A 0.1 in medium prewarmed to 42 °C (open symbols) or 30 °C (closed symbols) and incubation was continued. Samples were withdrawn every 5 (42 °C cultures) or every 15 (30 °C cultures) min for absorbance measurements and β -galactosidase assays^{35,37}. Strain construction is described in Table 1 legend.



response¹⁰⁻¹². In *recA* or *lexA* derivatives of the *sfiA::lacZ* fusion strain no UV-ray induction of β -galactosidase synthesis is observed in the dose range 1–10 J m⁻² (compare with Table 1). These same *recA* and *lexA* mutations introduced into a *gal::lacZ* fusion strain do not affect the inducibility of β -galactosidase synthesis by galactose, and the capacity of these highly UV-ray sensitive strains to synthesize β -galactosidase is not seriously reduced by 10 J m⁻² UV-ray irradiation (Table 1).

The *tif* and *tsl* mutations in the *recA* and *lexA* genes cause constitutive expression of some SOS functions at 42 °C (refs 18, 19), including *sfi*-dependent filamentation^{13,14,17}. At this temperature in *tif sfiA::lacZ* (Fig. 2a, Table 1) and *tsl sfiA::lacZ* (Fig. 2b, Table 1) fusion strains, a high rate of β -galactosidase synthesis is induced, but no similar induction is observed in a *tif gal::lacZ* fusion strain (Table 1). Furthermore, division inhibition by expression of the *tsl* mutations, unlike all other known inducing treatments, does not depend on *recA*⁺ activity²⁰ and in a *tsl recA sfiA::lacZ* fusion strain a high level of β -galactosidase synthesis is still observed at 42 °C (Fig. 2c, Table 1).

The *recA*⁺ and *lexA*⁺ gene products are thus regulators of *sfiA* gene expression, probably at the transcriptional level. The *recA* and *din* genes are similarly inducible in a *recA*⁺–*lexA*⁺–dependent manner^{19,21}, and in the case of the *recA* gene this regulation has been shown to be at the transcriptional level^{22,23}.

The *lon* mutation²⁴ amplifies *sfi*-dependent filamentation^{13,25,26}, sensitizing the cell to small doses of DNA damaging treatments. At least seven different enzymes involved in cell-wall biosynthesis are present at higher levels in *lon* strains than in *lon*⁺ strains²⁷, and in one case, that of the *gal* operon, the *lon* mutant has been reported to have increased transcription rates²⁸. The induction of *sfiA* gene expression by UV-ray irradiation was identical in *lon* and *lon*⁺ *sfiA::lacZ* fusion strains in basal level and induction kinetics (see Table 1). Thus *lon*⁺ function does not seem to be a transcriptional regulator of the *sfiA* operon. As *lon* mutants are also deficient in Deg protease activity²⁹, it is tempting to speculate that this protease is normally involved in eliminating excess *sfiA* protein, allowing rapid resumption of cell division as soon as normal DNA replication is restored.

Current models accounting for the regulation of the SOS response postulate that DNA damaging treatments result in the formation of a signal molecule by the DNA replication complex³⁰⁻³². This molecule activates the *recA* protein to a protease form which, in turn, cleaves certain cellular repressors, in particular the *cI* repressor of λ prophage³³ and the *lexA* protein³⁴. The results reported here show that a high rate of *sfiA* gene expression is part of the SOS response. The *sfiA* product is directly involved in SOS division inhibition and its synthesis is induced very rapidly during thymine starvation or after UV-ray

irradiation (Fig. 1). The *sfiA* product thus provides a highly sensitive link coupling cell division to DNA replication by the following probable sequence of events: perturbations or alterations of the DNA replication complex cause the generation of a signal molecule which activates the *recA* protein to its protease form. The *recA* protease then cleaves the *lexA* repressor, leading to derepression of the *sfiA* operon. Finally, the high concentration of *sfiA* protein thus produced results in an inhibition of cell division. In this sense the *sfiA* gene product can be considered an inducible division inhibitor, although its precise role in division inhibition remains to be elucidated.

We thank Maxime Schwartz for providing a *gal::lacZ* fusion strain and Micheline Jaques for helping in experimental work. This work was supported in part by grants ATP 4185 (CNRS-INRA) and ATP 72-79-104 (INSERM). This letter is dedicated to the memory of Jacqueline George, whose enthusiasm and insight initiated the research.

Received 4 November 1980; accepted 9 February 1981.

- Schaechter, M., Maaløe, O. & Kjeldgaard, N. O. *J. gen. Microbiol.* **19**, 592–606 (1958).
- Cooper, S. & Helmstetter, C. J. *molec. Biol.* **31**, 519–540 (1968).
- Howe, W. E. & Mount, D. W. *J. Bact.* **124**, 1113–1121 (1975).
- Helmstetter, C. E. & Pierucci, O. *J. Bact.* **95**, 1627–1633 (1968).
- Clark, D. J. *J. Bact.* **96**, 1214–1224 (1968).
- Donachie, W. D. *J. Bact.* **100**, 260–268 (1969).
- Jones, N. & Donachie, W. *Nature new Biol.* **243**, 100–103 (1973).
- Zaritsky, A. & Pritchard, R. *J. Bact.* **114**, 824–837 (1973).
- Witkin, E. M. *Proc. natn. Acad. Sci. U.S.A.* **57**, 1275–1279 (1967).
- Radman, M. in *Molecular Mechanisms of Repair of DNA* (eds Hanawalt, P. C. & Setlow, R. B.) 355–367 (Plenum, New York, 1975).
- Witkin, E. M. *Bact. Rev.* **40**, 869–907 (1976).
- Devoret, R. *Biochimie* **60**, 1135–1140 (1978).
- George, J., Castellazzi, M. & Buttin, G. *Molec. gen. Genet.* **140**, 309–332 (1975).
- McEntee, K. in *DNA Repair Mechanisms* Vol. 9 (eds Hanawalt, P. C., Friedberg, E. C. & Fox, C. F.) 349–359 (Academic, New York, 1978).
- Huisman, O., D'Ari, R. & George, J. *J. Bact.* **144**, 185–191 (1980).
- Casadaban, M. J. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4530–4533 (1979).
- Huisman, O., D'Ari, R. & George, J. *Molec. gen. Genet.* **177**, 629–636 (1980).
- Castellazzi, M., George, J. & Buttin, G. *Molec. gen. Genet.* **119**, 139–152 (1972).
- Gudas, L. J. *J. molec. Biol.* **104**, 567–587 (1976).
- Mount, D. W., Walker, A. C. & Kosel, C. J. *Bact.* **121**, 1203–1207 (1975).
- Kenyon, C. J. & Walker, G. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2819–2823 (1980).
- Gudas, L. J. & Pardee, A. B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2330–2334 (1975).
- McPartland, A., Green, L. & Echols, H. *Cell* **20**, 731–737 (1980).
- Howard-Flanders, P., Simson, E. & Theriot, L. *Genetics* **49**, 237–246 (1964).
- Gayda, R. C., Yamamoto, L. T. & Markovitz, A. J. *Bact.* **127**, 1208–1216 (1976).
- Johnson, B. F. *Genet. Res.* **30**, 273–286 (1977).
- Hua, S. & Markovitz, A. J. *Bact.* **110**, 1089–1099 (1972).
- Buchanan, C. E., Hua, S., Avni, H. & Markovitz, A. J. *Bact.* **114**, 891–893 (1973).
- Gottesman, S. & Zipser, D. *J. Bact.* **133**, 844–851 (1978).
- McEntee, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5275–5279 (1977).
- Gudas, L. J. & Mount, D. W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5280–5284 (1977).
- Emmerson, P. T. & West, S. C. *Molec. gen. Genet.* **155**, 77–85 (1977).
- Roberts, J. W., Roberts, C. W. & Craig, N. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4714–4718 (1978).
- Little, J. W., Edmiston, S. H., Pacelli, L. Z. & Mount, D. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3225–3229 (1980).
- Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor, New York, 1972).
- Kirby, E. P., Ruff, W. L. & Goldthwait, D. J. *Bact.* **111**, 447–453 (1972).
- Pardee, A. B., Jacob, F. & Monod, J. *J. molec. Biol.* **1**, 165–178 (1959).

Molecular cloning of foot and mouth disease virus genome and nucleotide sequences in the structural protein genes

John C. Boothroyd*, Peter E. Highfield*, George A. M. Cross*, David J. Rowlands†, Peter A. Lowe†, Fred Brown† & Timothy J. R. Harris†

* Department of Immunochemistry, Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, UK

† Department of Biochemistry, Animal Virus Research Institute, Pirbright, Woking, Surrey GU24 0NF, UK

Foot and mouth disease virus (FMDV), of the family Picornaviridae, consists of a single-stranded RNA (~8,000 nucleotides), the translation of which is initiated on the 3' side of a 150-nucleotide poly(C) tract and yields a single polypeptide which is processed by host cell proteases into four primary products (Fig. 1)¹⁻⁴. One or more virus-specified proteases further cleave these into the final products, the capsid proteins (VP1-4) being derived from the precursor p88 (for review see ref. 5). There are seven serotypes of the virus and as it has been shown that the immunizing activity of FMDV particles is associated primarily with VP1 (refs 6, 7), it seems likely that antigenic variation in FMDV is a result of changes in the structure of this protein. To further our understanding of this variation and as a first step in the possible development of FMDV vaccines from genetically manipulated microorganisms, we report here the construction and analysis of recombinant plasmids containing cDNA copies of the RNA. Comparison of the deduced amino acid sequence with the known polypeptide sequences shows that the NH₂-termini of VP2 and VP3 are conserved between the A and O serotypes whereas that of VP1 (the immunizing antigen) varies by as much as 42% between serotypes.

Using two strains of FMDV (serotypes A₁₀, strain 61 and O₁, strain BFS), double-stranded (ds) cDNA was synthesized⁸ and recombinant clones derived as described elsewhere⁹. Briefly, FMDV RNA (10 µg), primed with oligo (dT)₁₂₋₁₈, was used as a template for ds cDNA synthesis by reverse transcriptase. The ds cDNA so obtained ranged in size from 200 to 8,000 base pairs with a mean of ~1,000 base pairs as judged by agarose gel electrophoresis. The (dC)-tailed cDNA (0.05 pmol) was annealed with (dG)-tailed *Pst*I-digested pAT153 (0.2 pmol)¹⁰ and used to transform competent¹¹ *Escherichia coli* HB101 cells¹². Those clones derived from A₁₀ RNA are designated pFA10(61)/t whereas those from O₁ RNA are designated pF01(BFS)/t; these are abbreviated here to pFA and pFO, respectively. In both the pFA and pFO clone banks ~90% of the tetracycline-resistant colonies were sensitive to ampicillin (100 µg ml⁻¹) which suggested that these colonies carried plasmids with ds cDNA inserted at the *Pst*I site. Further characterization was initially on the basis of size, using a modification of the single-colony lysis procedure¹³. The size distribution paralleled that of the ds cDNA used in the cloning (data not shown); the largest plasmid, pFA206, contained an insert of 5,800 base pairs.

Construction of restriction maps of the four largest plasmids from both sets of recombinants showed that, in the case of the pFA set, two plasmids, pFA76 and pFA206, contained overlapping ds cDNA inserts which together corresponded to ~7,200 base pairs (~90% of the FMDV genome). In the pFO set, the largest plasmid, pFO251, contained all the sequence (~4,800 base pairs) present in the other pFO plasmids analysed. The *Pst*I site was not regenerated at both ends of all of the inserts which were analysed, probably because of slight exonuclease degradation at some stage during the construction of the recombinants.

To align the restriction maps with the physical map of FMDV, ³²P-labelled fragments of type A₁₀ RNA produced by T₁ RNase digestion were used as probes. Two oligonucleotides, T₁-3 and T₁-4, were used initially as these are located 1,000 (±300) and 2,300 (±500) nucleotides respectively from the 5' end of the RNA¹⁴. Aliquots of a *Pst*I/*Ava*I double digest of pFA76 were resolved by agarose gel electrophoresis, transferred to nitrocellulose filters¹⁵ and probed with ³²P end-labelled oligonucleotides. The two probes hybridized to different fragments of the

Fig. 1 Organization of the FMDV genome. The relative positions of the genes on the FMDV RNA have been published elsewhere⁵. VP1, 2, 3 and 4 are the capsid proteins and are derived from the precursor p88. The ds cDNA inserts of four recombinant plasmids are aligned with this map on the basis of restriction endonuclease analysis, hybridization to ³²P-labelled T₁ oligonucleotides (see Fig. 2) and on DNA sequence analysis. The T₁ oligonucleotides 20 and 27 are positioned on the basis of previously published results^{14,19}, whereas oligonucleotides 3 and 4 have been located exactly as a result of the DNA sequence analysis. The regions of DNA sequence presented in

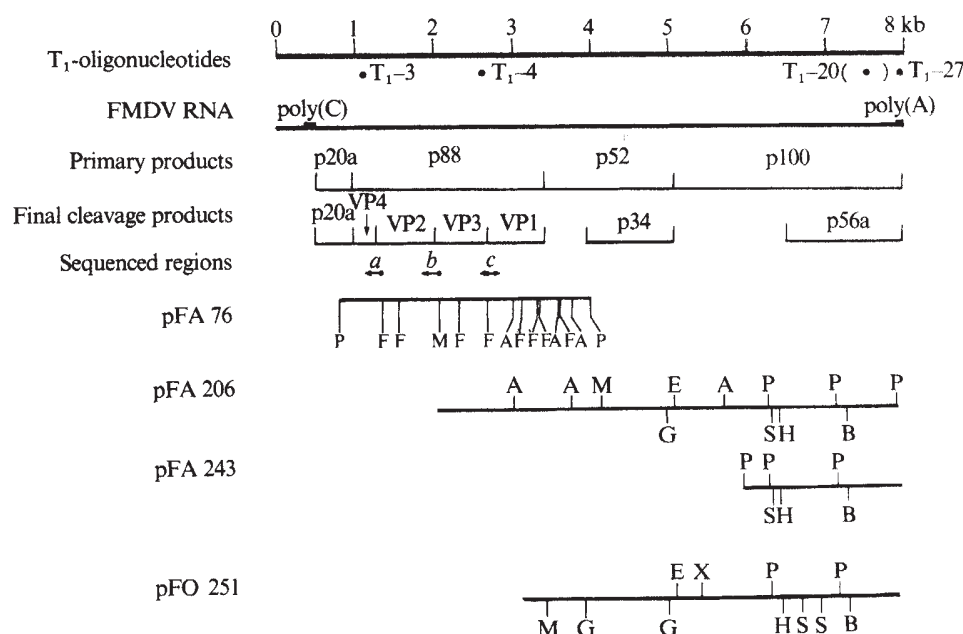


Fig. 3 are indicated by the heavy dots and arrowed lines over pFA76: a, b and c refer to the corresponding sections of Fig. 3. The restriction enzyme sites are designated: A, *Ava*I; B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; H, *Hind*III; M, *Sma*I; P, *Pst*I; S, *Sal*I; X, *Xba*I; F, *Hin*fI. The *Ava*I and *Hin*fI maps have been completely determined for pFA76 only.

Table 1 Homology between the amino termini of VP1 of five FMDV strains

	A ₁₂	O _{1K}	O _{1B}	C _{3R}
A ₁₀	69	82	58 ± 4	71 ± 4
A ₁₂		65	73 ± 4	83 ± 4
O _{1K}			63 ± 2	67 ± 4
O _{1B}				73 ± 6

Percentage homologies were determined from the data presented in Fig. 3. The potential error accounts for the ambiguity in the published sequences.

digest (Fig. 2) and, as the relative positions of these restriction fragments in the pFA76 ds cDNA insert were known, the inserted cDNA could be aligned with the RNA to within 500 nucleotides and its orientation in the plasmid obtained. The results indicated that the pFA76 insert spanned from ~800 to 4,000 nucleotides from the 5' end of the genome and the insert in pFA206, as deduced by its overlap with pFA76, spanned from ~2,000 to 7,800 nucleotides from the 5' end. This alignment was consistent with the finding that T₁-20 (which had been mapped to between 500 and 1,000 nucleotides from the poly(A) tail¹⁴) specifically hybridized to the *Pst*I fragment of pFA206 nearest the 3' end (data not shown). The ds cDNA contained within these two recombinants therefore included the entire region coding for the major capsid proteins (Fig. 1).

The pFA recombinant clones were also screened by colony hybridization^{16,17} to detect plasmids containing inserts derived from the ends of the RNA. Short, oligo(dT)-primed ³²P-cDNA was used as a probe for plasmids containing sequences from the 3' end of the RNA; several colonies gave a positive signal with this probe. The sensitivity of this assay was increased by isolating DNA from 1-ml cultures of these recombinant bacteria¹⁸ and binding this DNA onto nitrocellulose filters. Oligonucleotide T₁-27, which is located 12–21 nucleotides from the 3' end of the RNA^{14,19}, was ³²P-labelled and used to probe such a filter. Only one recombinant clone (pFA243) gave a positive signal; by DNA sequence analysis, this recombinant was found to contain sequences corresponding to the 3' end of the virus RNA, including ~30 nucleotides of the poly(A) tail (data not shown). From a comparison of restriction maps, pFA243 and pFA206 overlapped by ~2,000 base pairs (Fig. 1).

To detect plasmids containing sequences from the 5' end of RNA ³²P-pCp-labelled S fragment, the small (~380 nucleotides²⁰), 5'-terminal fragment of the RNA produced by RNase H digestion in the presence of oligo(dG)_{12–18} (ref. 3), was used to screen the recombinant clones. None of the clones gave a positive signal using this probe, which suggests that the 5' end of the RNA was not represented in the clone bank. In a similar hybridization with ³²P-labelled T₁-3 (located ~1,000 nucleotides from the 5' end, Fig. 1) only two recombinants (pFA76 and pFA187) gave positive signals (data not shown). Digests of pFA76 DNA were already known to hybridize with T₁-3 (Fig. 2, lane 3): further analysis of pFA187 by restriction endonuclease mapping showed that it extended only marginally further than pFA76 towards the 5' end (data not shown). These results indicate that ~750 nucleotides from the 5' end of the RNA are not represented in the recombinant clones.

An identical approach was used to orientate the pFO clones using the probes prepared from type A₁₀ RNA. This was possible even though there is only ~60% homology between the genomes of strains A61 and O₁BFS²¹. Radiolabelled T₁-20 specifically hybridized to one fragment of a *Pst*I/*Eco*RI double digest of pFO251 (Fig. 2, lane 5). The orientation of the ds cDNA insert relative to the genome suggested by this result is supported by the similarity between the restriction maps of pFO251 and pFA206 (Fig. 1).

A detailed restriction map of pFA76 was constructed and used to sequence defined parts within the region coding for the structural proteins (Fig. 1). The sequences obtained around the junctions of these proteins are presented in Fig. 3 together with

the predicted amino acid sequences. The latter are compared with the partial amino acid sequences which have been published for these proteins derived from other strains and serotypes of FMDV^{22–24}. The amino-terminal sequence of VP2 is identical in the A₁₀ and O_{1K} strains over the 20 amino acids available for comparison, while the amino terminus of VP3 shows just two differences out of 32 amino acids. The amino-terminal region of VP1 from A₁₀ can be compared with four other strains of FMDV (Fig. 3c). The relatively large extent of variation of the amino-terminal region of VP1 (compare with VP2 and VP3) may be related to the fact that this protein is the major immunizing antigen of FMDV^{6,7,25–27} and so variation of its sequence is under positive selection by the immune system of the host. The partial conservation of sequence which is apparent is probably due to constraints which must exist in order for VP1 to fulfill its function in the capsid and perhaps for the critical features of the protease cleavage site in p88 to be maintained. A comparison of the degree to which the sequences vary between strains (Table 1) shows no obvious correlation with the serological relationships of the different strains. However, as the region available for comparison constitutes only ~10% of VP1, it may be that the antigenic determinants which define its serotype lie elsewhere within the protein.

On the basis of their size and relative positions on the genome, it seems likely that the capsid proteins are contiguously coded²⁸.

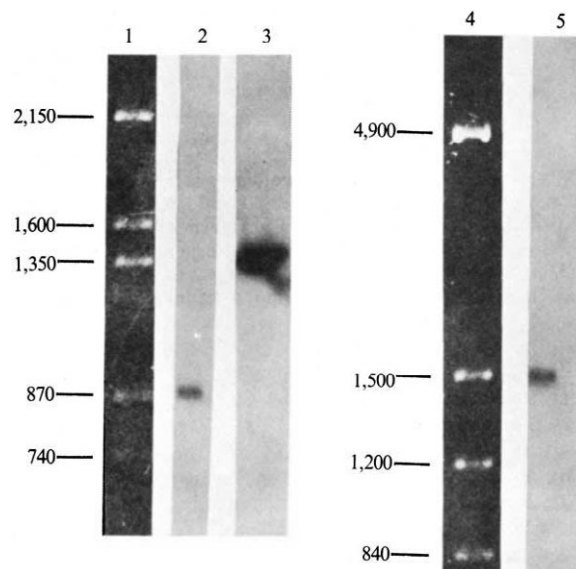


Fig. 2 Orientation of ds cDNA inserts with FMDV genome. Restriction fragments of pFA76 and pFO251 DNA were resolved by electrophoresis through a 0.85% agarose gel, and transferred to nitrocellulose filters (Schleicher & Schuell BA85)¹⁵. The filters were then incubated in 30 µl cm⁻² of hybridization buffer (40% (v/v) formamide, 0.4 M NaCl, 10 mM PIPES-NaOH pH 6.4, 0.5% (w/v) SDS, 0.04% (w/v) bovine serum albumin, 0.04% (w/v) Ficoll, 0.04% (w/v) polyvinyl pyrrolidone 40 and 100 µg ml⁻¹ *E. coli* tRNA) for 2 h at 50 °C before addition of the ³²P-labelled RNase T₁ oligonucleotide probes. The probes were prepared by two-dimensional gel electrophoresis of an RNase T₁ digest of 100 µg of ³²P-labelled type A₁₀ RNA (5 × 10⁵ c.p.m.)³. The oligonucleotides were eluted from the gel and labelled at the 5' end using polynucleotide kinase as described elsewhere³⁰. The probes were allowed to hybridize for 16–20 h at 50 °C before the filters were washed in 2 × SSC (eight changes of 250 ml each) at 45 °C, air-dried and autoradiographed for 24–28 h at -70 °C with an intensifying screen. Lane 1, ethidium bromide-stained gel of pFA76 DNA digested with *Pst*I and *Ava*I. Lane 2, parallel digest hybridized with ³²P-labelled T₁-4 (2.5 × 10⁵ c.p.m.). Lane 3, parallel digest hybridized with ³²P-labelled T₁-3 (4 × 10⁵ c.p.m.). Lane 4, ethidium bromide-stained gel of pFO251 DNA digested with *Pst*I and *Eco*RI. Lane 5, parallel digest hybridized with ³²P-labelled T₁-20 (5 × 10⁵ c.p.m.). The sizes of the restriction fragments in base pairs are indicated.

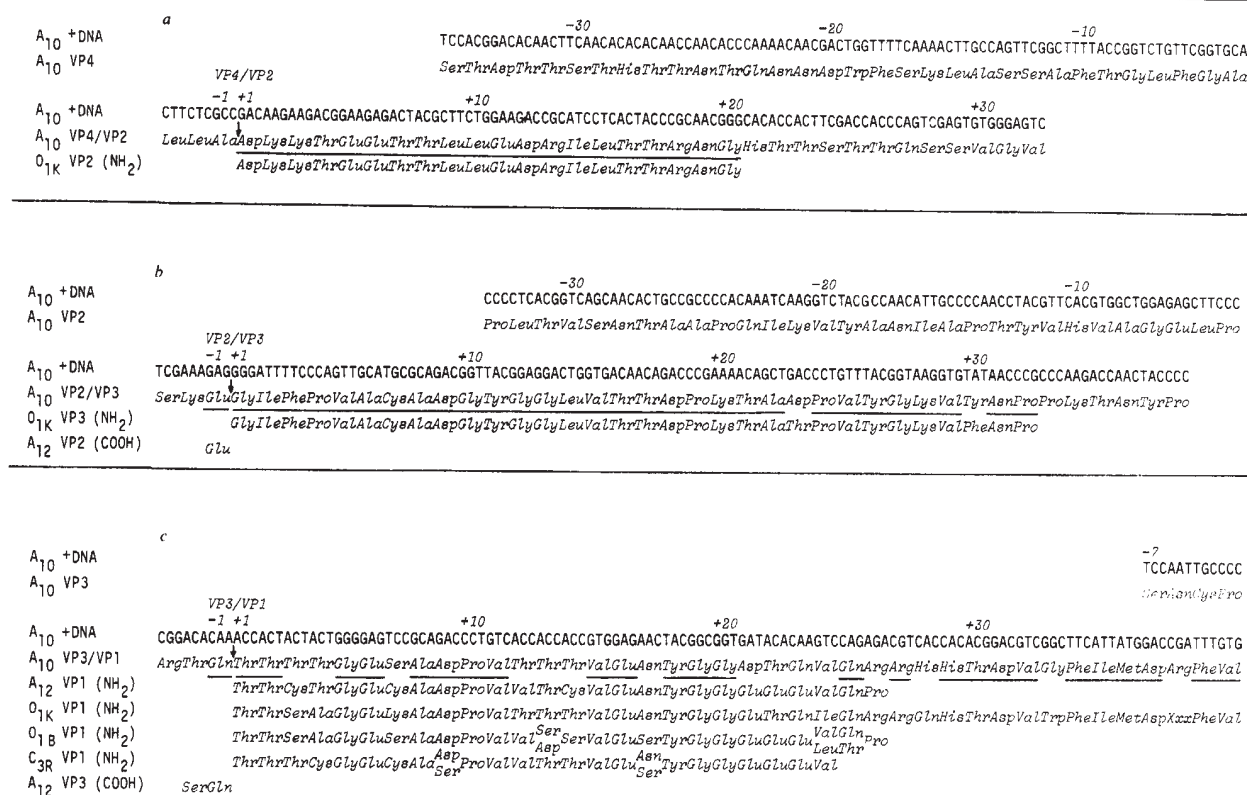


Fig. 3 Nucleotide and amino acid sequences at the junctions of the regions coding for the major capsid proteins. The nucleotide sequence of the positive strand of pFA76 was determined³¹ and used to deduce amino acid sequences in three regions (a, b and c; Fig. 1). The boundaries of the putative coding regions (↓) were set as described in the text. The amino acids are numbered from the amino-terminus of VP2, VP3 and VP1 in a, b and c, respectively. The published carboxy-terminal sequences (COOH) are from ref. 22; the published amino-terminal sequences (NH₂) from ref. 23 (type O_{1K}) and ref. 24 (types A₁₂, O_{1B} and C_{3R}). The underlined residues indicate homology between the A₁₀ sequence and the published data (ambiguities in the published amino-terminal sequences of VP1 are assumed to conform, where possible, to the consensus sequence).

The limited comparisons presented in Fig. 3 are consistent with this hypothesis: the sequence (Thr-Gln) which immediately precedes the amino terminus of VP1, is very similar to the published carboxy terminus of VP3 (Ser-Gln) from a different strain; the difference (Thr rather than Ser) is conservative with respect to charge and hydrophobicity. As the comparison for the carboxy terminus of VP2 is limited to only one residue (albeit identical with the published datum) and there are no comparative data for VP4, a definite conclusion on the contiguity of the structural polypeptides in p88 must await more extensive peptide analysis. Nevertheless, it is clear that the amino acid sequences at three proteolytic cleavage sites within p88 differ in both the type and identity of the amino acids.

VP1 purified from chemically disrupted virus particles has been shown to elicit low levels of neutralizing antibody in pigs and guinea pigs and to protect pigs against the disease²⁵⁻²⁷. An alternative approach to producing a vaccine therefore might be to prepare VP1 from genetically manipulated microorganisms. The recombinant pFA76 contains the sequence encoding all of p88 plus part of both p20 and p52. Such a plasmid offers the option of genetically engineering bacteria so that they synthesize all the capsid proteins in one precursor. This may have advantages over synthesis of VP1 by itself because in the normal replication cycle of FMDV, all capsid proteins are synthesized via the p88 precursor and this may influence their final conformations and immunological properties. Recombinant clones containing the coding sequences of all the capsid proteins, such as pFA76 described here, provide a starting point for such developments.

While this manuscript was in preparation we learned of similar work being done by Küpper *et al.*²⁹ on FMDV strain O_{1K}.

We thank Ruth Ashforth, Sharon Coleman and Cora Paynter for technical assistance and Drs J. Beard, R. B. Flavell, A. Newman, D. Sherratt and R. Thompson for gifts of materials and/or advice.

Received 1 December 1980; accepted 9 March 1981.

- Chatterjee, N. K., Bachrach, H. L. & Polatnick, J. J. *Virology* **69**, 369-377 (1976).
- Sangar, D. V., Rowlands, D. J., Harris, T. J. R. & Brown, F. *Nature* **268**, 648-650 (1977).
- Rowlands, D. J., Harris, T. J. R. & Brown, F. *Virology* **26**, 335-343 (1978).
- Sangar, D. V., Black, D. N., Rowlands, D. J., Harris, T. J. R. & Brown, F. *J. Virol.* **33**, 59-68 (1980).
- Sangar, D. V. *J. gen. Virol.* **45**, 1-13 (1979).
- Wild, T. F. & Brown, F. *J. gen. Virol.* **1**, 247-250 (1967).
- Wild, T. F., Burroughs, J. N. & Brown, F. *J. gen. Virol.* **4**, 313-320 (1969).
- Emlage, J. S., Catlin, G. H. & Carey, N. H. *Nucleic Acids Res.* **6**, 1221-1240 (1979).
- Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C. & Cross, G. A. M. *Gene* **8**, 391-417 (1980).
- Twigg, A. J. & Sherratt, D. *Nature* **283**, 216-218 (1980).
- Lederberg, J. & Cohen, S. N. *J. Bact.* **119**, 1072-1074 (1974).
- Boyer, H. W. & Roulland-Dussoix, D. *J. molec. Biol.* **41**, 459-472 (1969).
- Eckhardt, T. *Plasmid* **1**, 584-588 (1978).
- Harris, T. J. R., Robson, K. J. H. & Brown, F. *J. gen. Virol.* **50**, 403-418 (1980).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Hanahan, D. & Meselson, M. *Gene* **10**, 63-67 (1980).
- Grinstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961-3965 (1975).
- Birnboim, H. C. & Doty, J. *Nucleic Acids Res.* **7**, 1513-1523 (1980).
- Porter, A. G., Fellner, P., Black, D. N., Rowlands, D. J., Harris, T. J. R. & Brown, F. *Nature* **276**, 298-301 (1978).
- Harris, T. J. R. *J. Virol.* **36**, 659-664 (1980).
- Robson, K. J. H., Harris, T. J. R. & Brown, F. *J. gen. Virol.* **37**, 271-276 (1977).
- Bachrach, H. L., Swaney, J. B. & Van de Woude, G. F. *Virology* **52**, 520-528 (1973).
- Strohmaier, K., Wittman-Liebold, B. & Geissler, A. W. *Biochem. biophys. Res. Commun.* **85**, 1640-1645 (1978).
- Bachrach, H. L., Morgan, D. O. & Moore, D. M. *Intervirology* **12**, 65-72 (1979).
- Laporte, J., Grosclaude, J., Wantyghem, J., Serge, B. & Rouze, P. C. *r. hebdom. Acad. Sci. Séanc. Paris* **276**, 3399-3401 (1973).
- Meloen, R. H., Rowlands, D. J. & Brown, F. *J. gen. Virol.* **45**, 761-763 (1979).
- Bachrach, H. L., Moore, D. M., McKercher, P. D. & Polatnick, J. J. *Immun.* **115**, 1636-1641 (1975).
- Doel, T. R., Sangar, D. V., Rowlands, D. J. & Brown, F. *J. gen. Virol.* **41**, 395-404 (1978).
- Küpper, H. *et al. Nature* **289**, 555-559 (1981).
- Harris, T. J. R. *Nucleic Acids Res.* **7**, 1765-1785 (1979).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

BOOK REVIEWS

Newton unravelled

Marie Boas Hall

AS Voltaire remarked 250 years ago, few people read Newton because to do so intelligently one must be learned, but everybody talks about him. All scientists know a little of Newton but few know much, nor has it been easy to unravel the details of his life and work. For well over a century and a half scholars have known that Newton left a great mass of manuscript material; only over the past 30 years has it been studied seriously, transcribed and commented upon; and until now the standard biography has been that of David Brewster, published in 1855. Brewster had access to most of the manuscripts but though the Royal Society tried to make them accessible they were not finally deposited, in part, in the Cambridge University Library until 1888. No subsequent biographer utilized them, although Louis Trenchard More used and partly published some papers retained in private possession in his biography of 1934.

When Professor Westfall began work 20 years ago on this monumental scientific biography, destined to become the standard *Life* for the foreseeable future, he knew he must plunge into manuscript sources, for only Newton's *Correspondence* was in course of publication. As he worked, so the mathematical papers have been published *in extenso* by D.T. Whiteside, while many of the papers relating to the *Principia* and *Opticks*, early notebooks, thoughts on the nature of matter, selections from his historical, theological and alchemical manuscripts and numerous minor works have appeared with scholarly commentary, creating what Westfall, who has himself contributed to it, has called "the Newton industry".

Never at Rest is an able and thorough study of Newton the man and the scientist, especially and inevitably the latter who is both the more enduring and the better documented — almost too well documented. Newton was a compulsive scribbler: he read pen in hand, taking notes, making comments, writing and re-writing, analysing, commenting, organizing, drafting and redrafting, and seldom or never destroying.

Professor Westfall has worked through a great deal of the available manuscript material and presents a detailed account — careful, clear and judicious. First comes an analysis of the intellectual background of the day, then Newton's initiation into it,

Never at Rest: A Biography of Isaac Newton. By Richard S. Westfall. Pp.960. ISBN 0-521-23143-4. (Cambridge University Press: 1980.) £25, \$49.50.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

then his first steps towards original thought (in his reaction to Descartes' philosophical position) and his very first attempts at mathematics. Soon, leaping over conventional Euclidean geometry, Newton was deep into and going beyond contemporary developments in what we call analysis and he called "resolving problems by motion" (1665) and "the method of tangents", and "analysis by infinite series". In 1669, he showed a paper on the last subject (*De analysi*) to Isaac Barrow, first Lucasian Professor of Mathematics. Barrow immediately recognized its worth and when John Collins (a mathematical news monger, remembered for his

collection of contemporary letters published by Rigaud in 1841) sent him a new book which touched on series, Barrow in return sent *De analysi* (whence, in fact, news of the rising mathematician reached the Royal Society and through its secretary, Oldenburg, passed to Continental mathematicians). Had either Barrow or Collins been more adroit, *De analysi* would have been published and much later controversy avoided. When Barrow soon after resigned his professorship, Newton succeeded him; Westfall agrees with the tradition that Barrow's influence was here paramount.

Although Newton continued to work on and off at mathematics, he lectured on optics in which he had been interested for some years. The Latin lectures were deposited in the University Library as the regulations required, to be published only posthumously, but they became also Book I of the English *Opticks* in 1704, dealing with prismatic phenomena and dispersion. Soon Newton was led to the invention of the reflecting telescope, and of this, as he did not with mathematics, he wrote and spoke freely. The Royal Society learned of it in December 1671, elected him a Fellow, and with his permission both published an account and sent a written description to the Continent. Soon Oldenburg tactfully extracted Newton's first optical paper, with permission to publish it in the *Philosophical Transactions*. Although the Society praised the paper, it was by no means universally approved: Hooke, the Society's Curator of Experiments, an established scientist whose views Newton had earlier studied in *Micrographia* (1665), was acrimoniously and rudely critical; from Paris, Pardies (later a convert) and Huygens were politely so. Newton patiently answered their objections and in 1675, in the midst of controversy with the English Jesuits at Liège, sent his "Hypothesis on light and colours", whose publication he did not then permit. Yet he never blamed Oldenburg; indeed, after the latter's death in 1677, Newton told Hooke that he intended to publish his optical papers as Oldenburg had desired. Some sheets were printed, but the book never appeared — Westfall believes because of accidental destruction by fire.

Meanwhile Newton's interest was elsewhere — in chemistry or alchemy, an interest first aroused by his reading of

Mary Evans

Boyle, and then increasingly by the avid study of the vast literature of alchemy, on which he made incredibly extensive notes. He was an able experimenter, so much his notebooks reveal. There is no doubt that he thought alchemists rational, and believed that their strange language could be mastered. Evidently he believed that he understood alchemical symbolism; unfortunately he never translated it, and the key to his experimental programme is thus lost. Westfall believes that Newton's theory of attractive force was the outcome of this work — a debatable view. Newton never completed any papers on chemistry, and only one fragment was published in his lifetime, an incomplete theory of acids in the *Lexicon technicum* (1710) of John Harris. (Westfall unaccountably overlooked this publication, though he knew of other contemporary copies.) Many of Newton's views appear in drafts for a conclusion to the *Principia*, and were added to the *Queries* in *Opticks* in 1706 and 1710, in abbreviated form. At this time Newton also turned to theology, seeking for "primitive" Christianity and becoming a convinced, though secret, Arian.

But during these years he continued to work occasionally on mathematical and mechanical problems. Thus it is not very surprising that he could readily rise to Halley's challenge in 1684, to take up the problem of the Solar System in deep mathematical detail and produce, first a treatise *De motu* and then, under the Royal Society's encouraging praise and Halley's tactful prodding, the triumph of the *Principia* (1687). Westfall discusses the development of the work at some length, though there is more detail in the later chapter on Newton's revisions for the second edition (1713), when Newton was still working on major problems in old age, though less brilliantly.

With the publication of the *Principia*, whose importance was universally recognized, Newton's productive scientific life came near to an end. Perhaps, at 45, he knew it; at any rate, like many scientists he turned to other things. He was active in university affairs in 1688 and Member of Parliament for the University. He began to look for outside preferment, which he did not at once find. He suffered a brief mental breakdown in 1693. Westfall, totally rejecting Manuel's Freudian interpretation, for sound reasons, and equally the recent suggestion of mercury poisoning, ascribes it to a complex variety of factors, not least the mental exhaustion produced by the creative achievement of the *Principia*. Newton recovered, secured his job in London in 1696 as Warden (and later Master) of the Mint, and filled the posts with energy and distinction. In 1703 he became President of the Royal Society, here also showing both administrative ability and energetic dominance. Inevitably, he embarked on no new scientific work, although even in his seventies he was still capable of skilful application when

stimulated. He put *Opticks* into print, and revised it. He worked with Cotes on the second edition of the *Principia* from 1709 to 1713, while still busy at the Mint, and actively engaged in the vituperative dispute with Leibniz over the invention of the calculus. (Each refused to believe that the other had independently developed it, and each was driven on by hot-headed disciples.) The consequences of Newton's earlier failure to publish came home to roost. In extreme old age he turned again to theology, and to Biblical chronology; as with alchemy he sought reason and logic in mystical writings. He became the leading scientist of the first quarter of the eighteenth century, dominating optics, mechanics and mathematical astronomy; only the new mathematics owed nothing to him.

What of Newton the man? The biographical details are meticulously set out here as far as they exist: his family origins, his relations, his schooling, his life at Cambridge and in London. I find Westfall less successful in reconstructing Newton the man than Newton the scientist, whom he clearly prefers. Many of Newton's less agreeable characteristics were common in his day: his reluctance to publish and jealousy over priority were

shared by Huygens, as well as by many others. Was he really so isolated at Cambridge as tradition suggests? He got preferment early, and all his life he attracted and kept the loyalty of many young men. In the priority dispute he seems pathologically paranoiac; so does Leibniz; neither was attractive. In his old age Newton was clearly domineering, easily provoked to anger, impatient of lesser minds, vain and unjust to adversaries — he was never Voltaire's passionless hero. But neither was he totally and disagreeably ungenerous in his dealing with others, especially in his prime.

Never at Rest will be warmly welcomed by Newton scholars, although they will inevitably dispute some points. (Some may feel that Westfall in sticking so close to manuscript sources has been less than generous to earlier labourers.) The book is not easy going, but it provides a masterly, well-documented summary of contemporary views of all the many facets of Newton's astoundingly wide-ranging career which will be novel to all but the specialist, and will be essential reading for aspiring Newtonian scholars. □

Marie Boas Hall retired as Reader of History of Science and Technology at Imperial College, University of London, in 1980.

America and China: lessons from the past

Douglas P. Murray

American Science and Modern China 1876–1936. By Peter Buck. Pp.283. ISBN 0-521-22744-5. (Cambridge University Press: 1980.) £15, \$24.95.

THE recent and rapid embrace of Americans and Chinese has proceeded with little reflection on the troubled history of their cultural interactions. Inherent in this heavy but rewarding book is a call for caution lest we be fated "to repeat elsewhere the failure of American science in modern China". Not only elsewhere, but perhaps again in China itself.

Although Peter Buck describes his book as "an essay in comparative history", he is concerned far less with comparison than with the interaction of two societies that seemed destined to miss each other. He analyses the motives, problems and impact of those pioneers of modern science involved in the cross-cultural transmission of ideas and institutions during the late nineteenth and early twentieth century. Their failure is Buck's fundamental thesis, judged upon the stated goals of America's scientific missionaries and the view that modern science through the 1930s remained "wholly marginal to Chinese life". Yet in the sweep of his complex social history the criteria for what might have been success remain obscure; the legacy of American science that clearly is influencing

the People's Republic of China today is generally ignored.

As reflected in 25 pages of footnotes, the volume is meticulously researched, drawing liberally from, but not duplicating, the important work of other recent analysts of the West's educational impact in China. Buck's focus is somewhat different and original: the "transfer of scientific ideas and institutions to China primarily for what it shows about the United States, and only secondarily because of its bearing on Chinese scientific development" (p.1).

In a brief introduction, Buck assesses our own nineteenth-century debates about the proper role and social power of science and medicine, and his first chapter traces how these conceptions were developed and then acted out in China through medical missionaries. Chapters 3 and 4 focus, respectively, on the subsequent evolution of "laboratory science" (seen by one writer of the time as a "retreat . . . far removed from the strife of sects or wayward passions of the day") and its expression in the China Medical Board and Boxer Indemnity Scholarship program, and on the social origins and orientations of the US-trained scholars who in 1914 founded the Science Society of China.

Buck then returns to the United States to investigate the role of Cornell University

which by 1910 was, with Harvard, "enrolling more Chinese than any other American college" and, indeed, was the birthplace of the Science Society of China. As in the preceding chapter, the author-historian turns sociologist and here covers the most interesting ground. Viewing the development of American science in the context of a society undergoing industrialization, he demonstrates how Chinese were trained in the US for a world that would not exist for them at home and how "in formulating their plans for scientific development, the Science Society's leaders filtered out much of the social content with which American scientists had invested their images of research and laboratory science" (p.164). But, for Buck's purposes, the more critical point comes later, in Chapter 6: the result only further underscored how thoroughly interpretations of these ideals varied, even in the United States . . . In industrial America, Chinese scientists saw integration and order, whereas in reality there was intense conflict and turmoil . . . their image of American industrialization owed more to the ideals of American scientists than to the actual American experience (p.206). American conceptions of science tended to reach China after they already had changed or been challenged at home.

In a concluding chapter covering the 1920s and 1930s, Buck reviews the belated awakening of China's scientists to their dilemmas both in making science socially relevant and in coming to terms with the growing disappointments of their American benefactors. Yet, ironically, Chinese scholars also were becoming "progressively more convinced that the proper audience for their scientific work lay outside their country", with the international community (p.224). Making a rare link to more recent times, the author notes Mao Zedong's complaint about the behaviour of China's scientists before 1949:

He understood that they were not hostile or indifferent to the Chinese revolution, but from his perspective their enthusiasm for science involved a distinctive evasion of reality; [they mistakenly] 'stayed away from politics' (p.231).

Buck suggests a similar message for the West: detached from social context and public purpose, the best of science can be irrelevant. Our experience with China is the tapestry on which his message is woven.

In principle, this book should be required reading for all in the West now involved in or contemplating academic or technical cooperation with the PRC. Unfortunately, in practice it probably will appeal to a limited audience of scholars — those partial to the complex argumentation of an historian-philosopher whose commitment to grand concepts, intellectual rigour and direct quotation from a myriad of sources often makes the story hard to follow, or at least to enjoy. □

Douglas P. Murray is Director of the US China Relations Program at Stanford University.

Catching up with synchrotron radiation

Yves Farge

Synchrotron Radiation Research. Edited by Herman Winick and S. Doniach. Pp.754. ISBN 0-306-403-63-3. (Plenum: 1980.) £40.95, \$65.

THE past decade has seen an exponential growth in the use of synchrotron radiation — emitted by relativistic electrons in the bending magnets of storage rings — as an experimental probe. A consequence of this rapid advance is that, as in so many other fields, a book on the subject is already dated by the time it appears; this volume is no exception. Many of the 22 contributions will appear obsolete to the specialist, yet they do give a clear view of progress in synchrotron radiation research and of how its application is spreading from physics into chemistry, biology and applied sciences.

The first and last contributions are devoted to a discussion of how photons are generated by electrons passing through bending magnets. This radiation can range in wavelength from infrared to X-rays, is of very high intensity and its pulsed form can be used. Other advanced ideas, such as the use of wiggler or undulator magnets, which are now being tested on storage rings, and the development of a free-electron laser using undulators on these machines, are also discussed. In the remainder of the first part of the book, vacuum, ultraviolet and soft X-ray experiments on atoms, molecules and solids are considered; topics covered include absorption, fluorescence (radiative lifetimes and time-resolved spectroscopy) and photoelectron spectroscopy. These chapters give the reader a good view of the different kinds of electronic excitations which can be observed in such systems. One flaw, however, is that the chapter on photoelectron spectroscopy in solids is incomplete in that angle-resolved analysis of photoelectron spectra is largely ignored; even in 1978, from when most of the contributions date, this technique was known to be an important tool for band-structure determination in solids.

Of the later contributions, those on X-ray absorption spectroscopy are particularly good: Kronig's oscillations or EXAFS can provide useful structural information about a specific chemical element in inorganic as well as organic or biological samples. For example, study of the detailed structures of absorption edges reveals information about ionicity and coordination of the absorbing element.

The chapters on X-ray scattering provide an introduction to a field which was in its infancy at the time of writing (small-angle X-ray scattering, protein crystallography); they are mainly concerned with biological applications. Discussions of various applications, such as topography, X-ray fluorescence analysis, X-ray imaging and microscopy, complete the book.

This volume will be useful to non-specialists, as well as those already using synchrotron radiation in their research. One drawback is that work carried out at Stanford is emphasized at the expense of that performed in Europe, the USSR and Japan. However, the book clearly illustrates the revolutionary implications of the introduction of the technique into the natural sciences and is a cogent argument for the need for new machines designed specifically for the production of synchrotron radiation, some of which are under construction in several countries. □

Yves Farge was formerly Director of the French Synchrotron Radiation Facility (LURE), and is Chairman of the Synchrotron Radiation ad hoc Committee of the European Science Foundation and Scientific Counsellor at the Secrétariat d'État à la Recherche.

Reactions on a leash

Johannes Meienhofer

Polymer-Supported Reactions in Organic Synthesis. Edited by P. Hodge and D.C. Sherrington. Pp.484. ISBN 0471-277-12-6. (Wiley: 1980.) £29.50, \$88.

SOLID-PHASE synthesis of peptides and, recently, of oligonucleotides has received widespread attention, due in part to the interesting biological activities of these substances. Somewhat less notice has been given to the development of an impressive variety of polymer-supported organic chemical reactions in which a reagent, a catalyst or a substrate is attached to a polymer. There was a need to gather all aspects of polymer-supported organic synthesis into one volume, a task that the editors have accomplished admirably. A multitude of interesting reactions in many areas of chemistry is presented in 10 chapters, each written by an expert.

The opening contribution by D.C. Sherrington provides the reader (including the novice) with a solid base for understanding both the nature and the potentials of solid-support systems, in particular of the polystyrene type. P. Hodge then considers polymer-supported reagents. Oxidations, reductions and substitution reactions, condensing and acylating systems, solid phase ylids and organometallic reagents are presented. The large amounts of data are clearly organized in 42 well-designed tables and discussions of reaction mechanisms are aided by many schemes, equations and structures.

Sherrington discusses catalysis by ion-exchange resins and related materials, including a survey of transfer catalysis. Opportunities and constraints of catalysis

by enzyme-like polymers are covered by T. Kunitake, and those of transition metals by C.V. Pittman Jr. Chapter 6, on synthesis using polymer-supported protecting groups, by J.M.J. Fréchet, includes asymmetric synthesis and preparation of alkaloids, unsymmetrical carotenoids and other interesting syntheses. A sober and thorough assessment of the present scope and limitations of polymer-supported synthesis in general concludes this chapter. The contributions of J.M. Stewart, Fréchet and M.J. Gait describe solid-phase synthesis of peptides, oligosaccharides and oligonucleotides, respectively. Although the peptide and nucleotide topics have been reviewed frequently and in more detail elsewhere, these three skilfully prepared chapters are well worth reading. Solid-

phase oligosaccharide synthesis still needs much work to become competitive. Completing the main text is another chapter by Hodge, on various applications of polymer-supported species. Rounding off the book is an appendix which provides experimental guidance on the art of suspension polymerization, and on chemical modification of polymers.

This useful book gives thorough attention to detail and provides a wealth of clearly drawn illustrations. Research into polymer-supported reactions is a young and growing enterprise. This book should certainly further its development. □

Johannes Meienhofer is in the Chemical Research Department of Hoffmann-La Roche, Inc., Nutley, New Jersey, USA.

Of plants and pathogens

John Friend

Plant Disease: An Advanced Treatise. Vol. 5, How Plants Defend Themselves. Edited by James G. Horsfall and Ellis B. Cowling. Pp. 534. ISBN 12-356-4050. (Academic: 1980.) £27.80, \$49.50.

PLANTS have a range of natural defences which enable them to grow successfully and produce healthy crops despite the widespread presence of many potential pathogens. These resistance mechanisms have had an impact on plant evolution and further understanding of the processes is required in order that the mechanisms can be utilized and adapted by plant breeders and agronomists to improve crop production and yield.

In *How Plants Defend Themselves*, the final volume of a multi-author treatise on plant disease, the various ways in which plants resist disease are covered in 23 chapters. The early sections cover well-documented resistance phenomena such as escape from, and tolerance to, disease.

Later contributions deal with preformed barriers to attack, both physical and chemical. Physical barriers may be external, such as specialized types of plant surfaces and their coverings, or internal, including suberized and lignified tissues, gums, resins and silica. Preformed defensive chemicals include relatively simple compounds exuded from roots and specific antifungal and antibacterial compounds, often of chemotaxonomic interest, present inside the plants.

Considerable effort is now being expended on research into active mechanisms of plant resistance to pathogens, a topic which is covered in considerable depth in the book. Recognition and compatibility between host and parasite which are the most important initial events are surveyed in detail by Sequeira, who pays particular

attention to the possible functions of lectins, polysaccharides and lipopolysaccharides. Király reviews hypersensitivity with particular reference to his own recent experiments, and concludes that it is not necessarily the cause of resistance but probably a result of stress that develops in incompatible host-parasite combinations. Following a review of active formation of physical barriers by Beckman, Cruickshank provides a thoughtful review of the phytoalexin concept and the possible roles of phytoalexins in disease resistance.

However, the highlight of the book is a series of five chapters on acquired immunity to a second attempted invasion triggered by an initial infection by viruses, bacteria, fungi, nematodes and insects. In the main these chapters are descriptive, but the two on viruses and bacteria discuss possible mechanisms for this type of resistance which could involve alterations in nucleic acid and protein metabolism.

The phenomena are of great importance for an understanding of resistance mechanisms and may also have practical application in the design of new fungicides.

The book is an excellent summary of the most recent work on plant disease resistance. It can be criticized because it under-emphasizes the stimulus given by plant pathologists to research on the biochemistry and physiology of healthy plants; nevertheless, the reader should realize that the editors deliberately tried to produce a series in which the authors would write reflective articles examining the future of research on plant disease. Despite the rather anthropomorphic chapter headings and the idiosyncratic suggestion that plant resistance is analogous to the defence of a mediaeval castle, this editorial policy has been successful. □

John Friend is Head of the Department of Plant Biology at the University of Hull.

Nature's archives

Robert Olby

Natural History Manuscript Resources in the British Isles. Compiled by G.D.R. Bridson *et al.* Pp. 473. ISBN 0-7201-1559-0. (Mansell, London/Bowker, New York: 1980.) £97.

EVER since the establishment and consolidation of local government authorities in the nineteenth century, archive material has been preserved in provincial record offices, public libraries and civic universities on an increasing but uncoordinated basis. Successive parliamentary acts have expanded the list of official repositories for public documents, most of which are held either in a record office under the jurisdiction of the chief executive (formerly the town clerk) or in a public library under the control of the librarian. The scattering of archives between public libraries, record offices, museums, local societies, professional institutions and universities is inevitable given the varied sources generating them and the diversity of the parties interested in studying them. That the reform of local government in 1974 failed to clarify the relationship between the record office and public library is hardly surprising. That there should be some rivalry between civic universities and public libraries over the deposit of documents is to be expected. The important problem is that of guiding the inquirer in his search for source material.

There have been a number of initiatives to offer such guidance and to stimulate the formation of collaborative policy for deposition. The Commission on Historical Manuscripts, which celebrated its centenary in 1969, has played an invaluable role through the National Register of Archives which was established in 1949. Of particular importance for the history of science is the work of the Contemporary Scientific Archives Centre at Oxford.

The initiative for the present work goes back to the 1960s when D.E. Allen, R.H. Ellis and others formulated an ambitious plan to conduct a major search of the country's archives which relate to natural history. In the event adequate funds were not forthcoming and a more modest survey was undertaken by Bridson, Phillips and Harvey. They are to be congratulated for the extent and the detail of their book. There have always been close links between botany, zoology and natural history. Since the beginning of this century experimental biology has distanced itself from the work of the naturalists, but for most of the period covered by this survey students should find this book an aid to research. The only regret is that the authors were unable to include the extensive collections of the Public Record Office at Kew. □

Robert Olby is Reader in the History of Science, Department of Philosophy, University of Leeds.